

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 14:21:31 ; Search time 1590.15 Seconds
(without alignments)
4125.095 Million cell updates/sec

Title: US-09-743-690-6

Perfect score: 486

Sequence: 1 atgagcacaagtgtgtctca.....tgcgcacacagaagagtgga 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	378.2	77.8	538	10	BM489973	BM489973 ppg2n.pk0
2	244	50.2	460	9	AI980025	AI980025 pat.pk001
3	91.2	18.8	452	10	BG600657	BG600657 EST505552
4	91.2	18.8	457	10	B1920182	B1920182 EST540117
5	91.2	18.8	457	10	B922627	B922627 EST426396
6	89.6	18.4	452	10	B1432561	B1432561 EST535322
7	89.6	18.4	457	10	B923650	B923650 EST427419
8	89.6	18.4	458	10	BE472183	BE472183 EST417036
9	89.6	18.4	467	10	B1435772	B1435772 EST538533
10	88	18.1	330	10	BG600209	BG600209 EST505104
11	88	18.1	474	10	BG594050	BG594050 EST492728
12	88	18.1	475	10	B1178711	B1178711 EST519656
13	88	18.1	490	10	BG600210	BG600210 EST505105
14	87	17.9	455	10	BM406970	BM406970 EST581285
15	86.4	17.8	310	9	AW217469	AW217469 EST296183
16	86.4	17.8	378	10	B1432218	B1432218 EST534979
17	86.4	17.8	392	10	B1930261	B1930261 EST550150

18	86.4	17.8	463	10	BE353474	BE353474 EST533851
19	86.4	17.8	469	9	AI771798	AI771798 EST252898
20	86.4	17.8	464	10	BI431821	BI431821 EST534582
21	86.4	17.8	471	9	AI490242	AI490242 EST248568
22	86.4	17.8	493	9	AI486930	AI486930 EST245252
23	86.4	17.8	501	9	AW622766	AW622766 EST306752
24	86.4	17.8	520	10	BF460134	BF460134 066E06 Ma
c 25	84.8	17.4	692	9	AW616209	AW616209 EST307248
26	84	17.3	222	9	AW038839	AW038839 EST280795
27	84	17.3	460	10	BI435350	BI435350 EST538111
28	84	17.3	476	9	AW929928	AW929928 EST354198
c 30	83.2	17.1	486	9	AW092750	AW092750 EST285930
31	83.2	17.1	182	9	AJ319952	AJ319952 AJ319952
32	83.2	17.1	462	9	AI772725	AI772725 EST253825
33	83.2	17.1	469	9	AI774403	AI774403 EST255503
34	83.2	17.1	472	9	AI772154	AI772154 EST253254
35	83.2	17.1	475	10	BI433692	BI433692 EST536453
36	83.2	17.1	485	10	BI927178	BI927178 EST547067
37	83.2	17.1	489	9	AW040872	AW040872 EST283736
38	83.2	17.1	490	9	AW039971	AW039971 EST282462
39	83.2	17.1	490	9	AW039975	AW039975 EST282466
40	83.2	17.1	492	9	AW093509	AW093509 EST286689
41	82	16.8	487	10	BF153477	BF153477 027A01 Ma
42	81.6	16.8	203	10	BG628048	BG628048 CC-esf1cL
43	81.6	16.8	260	10	BE463103	BE463103 EST354242
44	81.6	16.8	471	9	AW039752	AW039752 EST282225
c 45	80.6	16.6	469	9	AI778944	AI778944 EST259823

ALIGNMENTS

BM489973 538 bp mRNA linear EST 07-FEB-2002
ppg2n.pk002.a9 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library (ppg2n) Gallus gallus cDNA clone ppg2n.pk002.a9 5' similar
to emb|CAC34569.1 (AJ311647) avidin [Gallus gallus], mRNA sequence.

BM489973

EST

chicken.

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 538)

Porter, F.E. and Cogburn, L.A.

ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA

Library, USDA/IFAFS Animal Genome Project

Unpublished (2002)

Contact: Larry A. Cogburn

University of Delaware

Townsend Hall, Newark, DE 19717, USA

Tel: 302-831-1335

Fax: 302-831-2822

Email: cogburn@udel.edu, www.chickest.udel.edu.

Location/Qualifiers

1. 538

/organism="Gallus gallus"

/strain="Commercial broiler chickens"

/db_xref="taxon:9031"

/clone="ppg2n.pk002.a9"

/clone.lib="Normalized Chicken

Pituitary/Hypothalamus/Pineal Library (ppg2n)"

/sex="Male and female"

/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"

/dev_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9

weeks)"

/lab_host="E. coli EMDH10B"

/note="Vector: pCMVSPORT6; Library made from equivalent

pools of total RNA isolated from each tissue at different

BASE COUNT 140 a 157 c 137 g 104 t
ORIGIN ages. Single pass sequencing from 5'-end"

Query Match 77.8%; Score 378.2; DB 10; Length 538;
Best Local Similarity 96.7%; Pred. No. 9, 1c-97;
Matches 386; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 88 ccagagatccctgcagaaagtgcgtgactgggaatggaccacacgacttgggctcc 147
Db 57 CCCGCCCTCTGTCGACAAAGCTGTCGCTGACTGCGAAATGGCAACACCTGGGCTCC 116
Qy 148 aacatgaccatcggggtgtgaacagagagtgatgaattcacaggcaactacacagcc 207
Db 117 AACATGACCATCGGGGTGTGAACAGGAAGGTGAATTCACAGGCACCTACACACAGCC 176
Qy 208 gtaacagccacataaatgagataaagagtcacattgcatggacacacacacac 267
Db 177 GTACAGCCACATCAATGAGATCAAGAGTCACCATGCGATGGGACACAAACACCATC 236
Qy 268 acaagaggagccagccacacattggcttcacggtcaattggagttttcagagtcaccc 327
Db 237 AACAGAGGCCACCCACCTTGGCTTCACTGTCAATTGGAAGTTCAGAGTCCAC 296
Qy 328 actgtcttcagggccagtgcttcatacagaggaatgggaaggaggtcctgaagaccatg 387
Db 297 ACTGTCTTCAGGCCAGTGTCTTATACAGAGGAAGGAGGAGGTCTTGAAGACCATG 356
Qy 388 tggctgtcggtcaagtgttaataagcattggtgatgactggaaggtaccagggtcggc 447
Db 357 TGCGTGTGGGTCAAGTGTAAATGACATTGTTGATGACTGGAAAGCTACAGGGTGGC 416
Qy 448 atcaacattctcactcctgcgcacacagagagtg 486
Db 417 ATCAACATCTTCACTCGCTCGGCACACAGAGAGTGA 455

RESULT 2
LOCUS AI980025 460 bp mRNA linear EST 07-MAY-2001
DEFINITION pat.pk0010.e3 chicken activated T cell cDNA Gallus gallus
clone pat.pk0010.e3 5', similar to avidin, mRNA sequence.
ACCESSION AI980025
VERSION AI980025.1 GI:5883053
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 460)
AUTHORS Tirunagaru, Y.G., Sofer, L., Cul, J. and Burnside, J.
TITLE An expressed sequence tag database of T-cell-enriched activated
chicken splenocytes: sequence analysis of 5251 clones
JOURNAL Genomics 66 (2), 144-151 (2000)
MEDLINE 20318616
COMMENT Contact: Joan Burnside
Molecular Endocrinology
University of Delaware
40 Townsend Hall, Newark, DE 19717, USA
Tel: 302 831-1345
Fax: 302-831-3411
Email: joan@udel.edu, www.chickest.udel.edu
Seq primer: 77.
Location/Qualifiers
1. .460
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pat.pk0010.e3"
/clone_lib="chicken activated T cell cDNA"
/sex="male"
/cell_type="Con A-activated splenic T cell"

FEATURES
SOURCE

BASE COUNT 119 a 118 c 111 g 89 t 23 others
ORIGIN /lab_host="E.coli top10 F-"
/note="Vector: pCDNA3"

Query Match 50.2%; Score 244; DB 9; Length 460;
Best Local Similarity 83.6%; Pred. No. 1, 3c-58;
Matches 280; Conservative 0; Mismatches 50; Indels 5; Gaps 1;
Qy 88 ccagagatccctgcagaaagtgcgtgactgggaatggaccacacgacttgggctcc 147
Db 84 CCCGCCCTCTGTCGACAAAGCTGTCGCTGACTGCGAAATGGCAACACATCGGCTCC 143
Qy 148 aacatgaccatcggggtgtgaacagagagtgatgaattcacaggcaactacacagcc 207
Db 144 AACATGACCATCGGGCTGTGAACAGCAGAGGTGAATTCACAGGCACCTACATCAGCC 203
Qy 208 gtaacagccacataaatgagatcaaaagagtcacattgcatgggacacacacacac 267
Db 204 GTAACAGCCACATCAATGAGATCAAGAGTCACCATGCGATGGGACACAAACACCATC 263
Qy 268 acaagaggagccagccacacattggcttcacggtcaattggagttttcagagtcaccc 327
Db 264 AACAGAGGGGCCAGCCACCTTGGCTTCACTGCGAAATGGGAAGTTCAGAGTCCACC 323
Qy 328 actgtcttcacgggcccagtgcttcatacagaggaatgggaaggaggtcctgaagaccatg 387
Db 324 ACTGTCTTACGGCGCAATCTTATANANGGAAATGGGAAGGGNNTTAAGACCA----- 378
Qy 388 tggctgtcggtcgaagtgttaatagacattggtga 422
Db 379 TTNGTCTCTCGGGCAATGTTAANNATTGGGTGA 413

RESULT 3

LOCUS BG600657 452 bp mRNA linear EST 12-APR-2001
DEFINITION EST505552 csts solanum tuberosum cDNA clone CTS29N13 5' sequence,
mRNA sequence.
ACCESSION BG600657
VERSION BG600657.1 GI:13617793
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 452)
AUTHORS van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A.,
Bougri, O., Buell, C.R., Ronning, C., Tanksley, S., and Baker, B.
TITLE Generations of ESTs from sprouting potato eyes
JOURNAL Unpublished (2000)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
Location/Qualifiers
1. .452
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CTS29N13"
/clone_lib="CSTS"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pbluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from

FEATURES
source

infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from *Phytophthora* rather than potato."

BASE COUNT	159 a	67 c	90 g	151 t	
ORIGIN					
Query Match	18.4%;	Score 89.6;	DB 10;	Length 467;	
Best Local Similarity	95.8%;	Pred. No. 1e-14;			
Matches	92;	Conservative	0;	Mismatches	4;
				Indels	0;
				Gaps	0;

QY	1	atggagtc	aaagtgtgctcacatcattgtttctttctttctttgcaactccctttgaaact	60
Db	7	ACGGAGTCAAAGTTGCTCACATCATGTGTTCTTCTTCTTGTGCAACTTCCTTTGAAACT	66	

QY	61	ctcttgacgacgaaagaaagtgtgaccagagatc	96	
Db	67	CTCTTGGCACGAAAGAAAGTGTGACCAAGATC	102	

RESULT	10				
BG600209					
LOCUS	EST505104	cSTS Solanum tuberosum	cdNA clone cST52818	5'	sequence,
DEFINITION	mRNA sequence.				
ACCESSION	BG600209		330 bp	mRNA	linear
VERSION	EST.				EST 12-APR-2001
KEYWORDS	BG600209.1	GI:13617345			
SOURCE	potato.				
ORGANISM	Solanum tuberosum				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.				
AUTHORS	1 (bases 1 to 330)				
TITLE	von der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chieningo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B. Generations of ESTs from sprouting potato eyes				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com Seq primer: M13F-R.				

FEATURES	Location/Qualifiers				
Source	1..330				
	/organism="Solanum tuberosum"				
	/cultivar="Kennebec"				
	/db_xref="taxon:4113"				
	/clone="cST52818"				
	/clone_lib="cSTS"				
	/tissue_type="sprouting eyes from tubers"				
	/dev_stage="12-14 weeks post harvest"				
	/lab_host="SOL"				
	note=Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."				

BASE COUNT	107 a	56 c	63 g	104 t	
ORIGIN					
Query Match	18.1%;	Score 88;	DB 10;	Length 330;	
Best Local Similarity	94.8%;	Pred. No. 2.7e-14;			
Matches	91;	Conservative	0;	Mismatches	5;
				Indels	0;
				Gaps	0;

QY	1	atggagtc	aaagtgtgctcacatcattgtttctttctttctttgcaactccctttgaaact	60
Db	35	ATGGAGTTAAAGTTGCTCACATCATGTGTTCTTCTTCTTGTGCAACTTCCTTTGAAACT	94	

QY	61	ctcttgacgacgaaagaaagtgtgaccagagatc	96	
----	----	------------------------------------	----	--

COMMENT

Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES

```

Location/Qualifiers
1. 490
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="CSTS28110"
/clone_lib="cSNS"
/tissue_type="sprouting eyes"
/dev_host="12-14 weeks post
/lab_host="SOLR"
/notes="Vector: pBluescript SK
xhoI; Various sizes of sprout
taken from tubers. The tubers
dark for 2-3 weeks prior to s
frozen in liquid nitrogen imm
tubers."
72 c 93 q 165 t
a

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Labels.	72 c	93 q	165 t
a			

Query

Query Match 18.1%; Score 88; DB 10; Length 490;
Best Local Similarity 94.8%; Pred. NO. 3e-14;
Matches 91; Conservative 0; Mismatches 5; Indels

Qy

Qy . 1 atggagtc aaagt tigtctcacatcatgttttcttctccttgcaactccctttgaaact 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 35 ATGGAGTTTAAAGTTTGCTCACATCATGTGTTTTCAATTCTTCCTTGCAACTCCTTTGAAACT 94

Oy

Qy 61 ctcttggcacgaaaagaagtgtatggaccagagatc 96
|||||
Db 95 CTCTTGGCACGAAAAGAATGTGATGGACCAGAAGTC 130
|||||

RESULT 14
BM406970

	BM406970	455 bp	mRNA	linear	EST 22-JAN-2002
LOCUS	EST581285	potato roots	Solanum tuberosum	cDNA clone	CPR029K4 5' end
DEFINITION	, mRNA sequence.				

ACCESSION

ACCESSION	BM406970
VERSION	BM406970.1
KEYWORDS	EST.
SOURCE	potato.

ORGAN:

ORGANISM

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE

REFERENCE
AUTHORS

TITLE

JOURNAL COMMENT
Unpublished (2001)
Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195

For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: T3.

FEATURES

```
1. 455
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cPRO29K4"
/clone_lib="potato roots"
/tissue_type="roots"
/dev_stage="in vitro grown st
/lab_host="SOLR"
```

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Cornell University, Tanksley lab; sequencing: The Institute for Genomic Research. Roots were isolated from in vitro grown stem cuttings on CM medium. Roots were isolated two weeks after placing the stem cuttings from in vitro grown plants on medium."

BASE COUNT	146 a	66 c	89 g	154 t
ORIGIN				

Query Match 17.9%; Score 87; DB 10; Length 455;
Best Local Similarity 94.7%; Pred. No. 5.6e-14;
Matches 90; Conservative 0; Mismatches 5; Indels

Qy	2	tgga	gtaaa	gttg	gtctacac	atcatt	gtttcttctctctctgcgaactccctttgaaactc	61
Db	1	TGGAGT	CAAAAGTTGGCT	CACATCAT	TGT	TTCTTTCTTCTTCTGCCACTTCCCTTTGAAACTC	60	
Qy	62	tcttggc	acgaaaagaa	agtgatg	agccacagagatc		96	
Db	61	TC	TTGGGACGACGAAAAGAA	TGTGATGGACACAGAGTC			95	

RESULT 15

LOCUS	AW2117469	310 bp	mRNA	linear	EST 18-MAY-2001
DEFINITION	EST296183 tomato flower buds 3-8 mm, Cornell University Lycopersicon esculentum cDNA clone cTOB1A15 similar to tomato leaf wound-induced proteinase inhibitor I, mRNA sequence.				

WOUND-INDUCED PROTEIN
ACCESSION AW217469

RECESSION	AW217469	GI:6528343
VERSION	AW217469.1	
KEYWORDS	EST.	
SOURCE	tomato.	

SOURCE
ORGANISM

Eucorpsicon eucorpsicum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE

AUTHORS van der Hoeven, R.S., Bezzerides, J.I., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.

TITLE Generation of ESTs from tomato flower tissue

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI

COMMENT

Correspondence: 2002
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES

```

1. 310
source
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOBLA15"
/clone.lib="tomato flower buds 3-8
/tissue_type="flower"
/dev_stage="3-8mm buds"

```

BASE COUNT

ORIGIN

Query Match	17.8%	Score	86.4	DB	9	Length	310
Best Local Similarity	93.8%	Pred. No.	7.5e-14				

Qy 1 atqgaqtcaaaqtttgctcacatcattgtttttctttcttgcactccctttgaaact 60

Search completed: September 20, 2002, 15:50:02
Job time: 5311 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 15:19:26 ; Search time 41.22 Seconds
(without alignments)
2896.116 Million cell updates/sec

Title: US-09-743-690-6
Perfect score: 486
Sequence: 1 atggagtcacaaagtgtgtcga.....tgcgcacacacagaagtgga 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	389.4	80.1	604	4 US-08-831-399-3	Sequence 3, Appli
2	279.8	57.6	484	1 US-08-554-586-1	Sequence 1, Appli
c 3	37	7.6	7218	1 US-08-232-463-14	Sequence 14, Appl
4	31.4	6.5	8040	1 US-08-596-291-1	Sequence 1, Appli
5	31.4	6.5	8040	3 US-09-100-804-1	Sequence 1, Appli
6	31.4	6.5	8043	5 PCT-US94-09943-1	Sequence 1, Appli
7	31	6.4	1662	2 US-08-956-012-2	Sequence 2, Appli
c 8	30.8	6.3	954	1 US-08-170-294-1	Sequence 1, Appli
c 9	30.8	6.3	954	2 US-08-664-855-1	Sequence 1, Appli
c 10	30.8	6.3	954	3 US-09-049-289-1	Sequence 1, Appli
c 11	30.6	6.3	400	4 US-09-022-238-1	Sequence 1, Appli
c 12	30.6	6.3	1644	1 US-07-675-211-1	Sequence 1, Appli
c 13	30.6	6.3	1644	1 US-07-903-047-1	Sequence 1, Appli
c 14	30.6	6.3	1644	1 US-08-076-042-1	Sequence 1, Appli
c 15	30.6	6.3	1644	2 US-08-757-046A-3	Sequence 3, Appli
c 16	30.6	6.3	1644	3 US-09-447-208-3	Sequence 3, Appli
c 17	30.6	6.3	1644	4 US-09-135-988-3	Sequence 3, Appli
c 18	30.6	6.3	1644	4 US-09-277-716-3	Sequence 3, Appli
c 19	30.6	6.3	1644	4 US-08-597-274A-3	Sequence 3, Appli
c 20	30.6	6.3	1644	4 US-09-380-061B-13	Sequence 13, Appl
c 21	30.6	6.3	1656	3 US-09-111-752-6	Sequence 6, Appli
c 22	30.6	6.3	1656	3 US-09-111-752-8	Sequence 8, Appli
c 23	30.6	6.3	1656	3 US-09-111-752-9	Sequence 9, Appli
c 24	30.6	6.3	1779	4 US-09-371-696-1	Sequence 1, Appli
c 25	30.6	6.3	3410	4 US-09-020-956-110	Sequence 110, App
26	30.6	6.3	3410	4 US-09-030-607-110	Sequence 110, App
27	30.6	6.3	3410	4 US-09-439-313-110	Sequence 110, App

c 28	30.4	6.3	2280	4 US-09-051-696-1	Sequence 1, Appli
29	30.2	6.2	3969	1 US-08-026-138E-16	Sequence 16, Appli
30	30.2	6.2	11707	4 US-09-136-574A-1	Sequence 1, Appli
31	30	6.2	1320	1 US-08-176-412-1	Sequence 1, Appli
32	30	6.2	1320	2 US-08-555-268A-1	Sequence 1, Appli
33	30	6.2	1320	2 US-08-495-695B-1	Sequence 1, Appli
34	30	6.2	1320	5 PCT-US94-14436-1	Sequence 1, Appli
c 35	29.8	6.1	1617	3 US-08-834-655-1	Sequence 1, Appli
c 36	29.8	6.1	1617	3 US-08-834-033A-1	Sequence 1, Appli
c 37	29.8	6.1	1617	3 US-09-363-574-1	Sequence 1, Appli
38	29.8	6.1	1830	1 US-08-410-804-2	Sequence 2, Appli
39	29.8	6.1	1830	1 US-08-259-514-2	Sequence 2, Appli
40	29.8	6.1	1830	2 US-08-858-311-2	Sequence 2, Appli
41	29.6	6.1	8119	4 US-09-290-640-45	Sequence 45, Appli
c 42	29.6	6.1	3870	1 US-08-138-641-1	Sequence 1, Appli
c 43	29.6	6.1	3870	1 US-08-138-133-1	Sequence 1, Appli
c 44	29.6	6.1	3893	1 US-08-138-641-3	Sequence 3, Appli
c 45	29.6	6.1	3893	1 US-08-138-133-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-831-399-3
; Sequence 3, Application US/08831399
; Patent No. 6312916
; GENERAL INFORMATION:
; APPLICANT: Kopezki, Erhard; Muller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Hans
; TITLE OF INVENTION: Recombinant Inactive Core
; TITLE OF INVENTION: Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831.399
; FILING DATE: 1-April-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 37 718.8
; FILING DATE: 16-September-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6312916man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 1105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 44..499
; OTHER INFORMATION: Positions 44..115 correspond to sig
; OTHER INFORMATION: peptide and 116..499 to mat peptide

APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 7.6%; Score 37; DB 1; Length 7218;
Best Local Similarity 4.1%; Pred. No. 0.055;
Matches 16; Conservative 203; Mismatches 168; Indels 0; Gaps 0;
Qy 66 ggcacgaaagaagtgtgacagagatccctgcagaaagtgtcgtgactggaa 125
Db 1426 RR 1367
Qy 126 atggaccaacgtctggtccaaacatgaccatcggtgtgaaacagagagtgatt 185
Db 1366 RR 1307
Qy 186 cacaggcactacatcacagcgttaacagccacatcaaatgagatcaaaagagtcaccatt 245
Db 1306 RR 1247
Qy 246 gcatgggacacaaacacatcaacagagagccagccaccttggcttcaccgtcaa 305
Db 1246 RR 1187
Qy 306 ttggaagtgttcagagtcaccactgtcttcacggtccaggtcttcacagaggaatgg 365
Db 1186 RR 1127
Qy 366 gaagaggtcctgaagaccatgtgctgtcggtcaagtgttaatgacattggtgatga 425
Db 1126 RR 1067
Qy 426 ctggaagctaccagggtcgcatcaa 452
Db 1066 ATCGCAAGCTCCCTGACCTGCAGCCA 1040

RESULT 4
US-08-596-291-1
Sequence 1, Application US/08596291
Patent No. 5821075
GENERAL INFORMATION:
APPLICANT: GONEZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESSON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA

ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,291
FILING DATE: 09-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: L0461/7000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8040 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
FEATURE:

NAME/KEY: CDS
LOCATION: 78..7475
US-08-596-291-1
Query Match 6.5%; Score 31.4; DB 1; Length 8040;
Best Local Similarity 57.7%; Pred. No. 4;
Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
Qy 170 acagcagaggtgaaattcacagcagcaccatcacagccgttaacagccacatcaaatgaga 229
Db 4978 ACACGAGTGGAGTGGAGAGATGACTTAGTCACAGCTCCACCAACATATCAAAATCCA 5037
Qy 230 tcaagagtcaccattgcatgggacacaaacacacatt 266
Db 5038 CTTGGAGTTCAGCTTTGTCATCAGACTCTAAGCAACAT 5074

RESULT 5
US-09-100-804-1
Sequence 1, Application US/09100804
Patent No. 6066472
GENERAL INFORMATION:
APPLICANT: GONEZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESSON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/664,855
FILING DATE: 17-JUN-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/170,294
FILING DATE: 30-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9114259.6

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 15:22:41 ; Search time 178.23 Seconds
(without alignments)
4681.706 Million cell updates/sec

Title: us-09-743-690-6
Perfect score: 486
Sequence: 1 atggagtcaaaattgtctca.....tcgcacacagaagagtga 486

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	486	100.0	486	21	AAZ49865
2	394.2	81.1	576	21	AAZ49855
3	389.4	80.1	459	20	AAZ498197
4	389.4	80.1	604	18	AAZ73194
5	386.4	79.5	5177	20	AAZ09997
6	279.8	57.6	484	18	AAZ68502
7	94.4	19.4	324	21	AAZ49854
8	36.6	7.5	1700	23	AAZ85309
9	34.8	7.2	2781	23	ABL08242

10	34.8	7.2	4512	23	ABL08255	Drosophila melanog
11	34.8	7.2	6585	23	ABL08254	Drosophila melanog
12	34.6	7.1	2385	13	AAQ27956	Cholesterol oxidas
13	33.8	7.0	1784	20	AAZ58334	Mouse macrophage-d
14	33.8	7.0	1944	19	AAV34998	Mouse chemokine pr
15	33.4	6.9	2543	21	AAZ52774	Soybean putative c
16	33.4	6.9	9436	22	AAZ18294	Mouse massl cDNA m
17	33.4	6.9	9437	22	AAZ18286	Mouse monogenic au
18	33.2	6.8	1444	22	AAH33104	Human colon cancer
19	33.2	6.8	2501	22	AAH76844	Human regulatory t
20	33	6.8	1341	23	ABU26493	Drosophila melanog
21	33	6.8	3042	23	ABU26492	Drosophila melanog
22	32.6	6.7	8170	22	AAU07459	Human reproductive
23	32.6	6.7	8170	22	AAU07461	Human reproductive
24	32.6	6.7	9993	22	AAU07460	Human reproductive
25	32.4	6.7	1503	22	AAK69825	Human immune/haema
26	32.4	6.7	1503	22	AAK69827	Human immune/haema
27	32.4	6.7	1503	22	AAK69829	Human immune/haema
28	32	6.6	1355	21	AAZ69568	Human secreted pro
29	32	6.6	2312	22	AAI59647	Human polynucleoti
30	32	6.6	2376	22	AAI57861	Human polynucleoti
31	32	6.6	2629	9	AAH1338	Protein A gene wit
32	32	6.6	2629	13	AAQ24523	Vasoactive intesti
33	32	6.6	8436	23	ABU10730	Drosophila melanog
34	32	6.6	513445	22	AAI61373	Soybean 318013 reg
35	31.8	6.5	46593	21	AAZ81456	N. meningitidis pa
36	31.8	6.5	349980	21	AAZ21612	Neisseria meningit
37	31.8	6.5	837096	21	AAZ17107	N. meningitidis pa
38	31.6	6.5	834	23	AAZ71707	DNA encoding novel
39	31.6	6.5	1414	23	AAZ79889	DNA encoding novel
40	31.6	6.5	14041	22	AAH48024	Internal control B
41	31.4	6.5	677	23	ABU07791	Drosophila melanog
42	31.4	6.5	2544	22	AAZ94367	Human full-length
43	31.4	6.5	2799	23	ABU07790	Drosophila melanog
44	31.4	6.5	8043	16	AAQ85924	Human protein tyro
45	31.4	6.5	112190	22	AAH44801	Human GPCR protein

ALIGNMENTS

RESULT	1
AAZ49865	
ID	AAZ49865 standard; cDNA; 486 BP.
AC	AAZ49865;
XX	
XX	
DT	25-APR-2000 (first entry)
XX	
DE	Potato proteinase inhibitor-I/Avidin gene fusion sequence.
XX	
KW	Potato proteinase inhibitor-I; PPI-I; streptavidin; worm;
KW	insect; plant-noxious protein; pest resistance; moth; insect; weevil;
KW	grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;
KW	insecticidal; ss.
OS	Solanum tuberosum.
OS	Synthetic.
PH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..486
FT	/*tag= a
FT	/product= "PPI-I/Avidin fusion protein"
XX	
PN	WO200004049-A1.
XX	
PD	27-JAN-2000.
XX	
PF	15-JUL-1999; 99WO-NZ00110.
XX	
PR	15-JUL-1998; 98NZ-0331002.
XX	
PA	(HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.


```

Db 104 cccggatccctgcagaaagtgctcgtgactggaatggacaaacgactcgggctcc 163
QY 148 aacatgaccatcgggcgtgtgaacagcagaggtgaattccacagggcacctacacacagcc 207
Db 164 aacatgaccatcgggcgtgtgaacagcagaggtgaattccacagggcacctacacacagcc 223
QY 208 gtaacagccacatcaaatgagatcaaaagagtcaccattgcatggagacacaaacaccatc 267
Db 224 gtaacagccacatcaaatgagatcaaaagagtcaccattgcatggagacacaaacaccatc 283
QY 268 aacaagagagaccagcccaccttggcttcacagtcacaaattggaagttttcagagtcacc 327
Db 284 aacaagagagaccagcccaccttggcttcacagtcacaaattggaagttttcagagtcacc 343
QY 328 actgtcttcacggccagtgcttcacagcaggaatgggaagaggtcctgaagacacatg 387
Db 344 actgtcttcacggccagtgcttcacagcaggaatgggaagaggtcctgaagacacatg 403
QY 388 tggctgctcggtccaagtgttaatgacattggtgatgactggaaggtaccagggctcgcc 447
Db 404 tggctgctcggtccaagtgttaatgacattggtgatgactggaaggtaccagggctcgcc 463
QY 448 atcaaatcttcaactgcctgcgcacacagagaggtga 486
Db 464 atcaaatcttcaactgcctgcgcacacagagaggtga 502

RESULT 3
AAX80197
ID AAX80197 standard; cDNA; 459 BP.
AC AAX80197;
XX
XX
PT 20-AUG-1999 (first entry)
DE
DE Chicken avidin encoding cDNA.
KW Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;
KW hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin; ss.
XX
OS Gallus gallus.
XX
XX WO929838-A1.
XX
XX 17-JUN-1999.
XX
XX 09-DEC-1998; 98WO-US26086.
XX
XX 09-DEC-1997; 97US-0067978.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Cedarholm-Williams SA;
XX
XX WPI; 1999-385599/32.
XX
XX P-PSDB; AAY17867.
XX
XX A fibrinogen-converting enzyme fusion protein
XX
XX Disclosure; Page 26-27; 35pp; English.
XX
XX The present invention describes a fibrinogen-converting enzyme fusion
XX protein (FCE). The fusion protein is a multidomain protein comprising:
XX (a) a FCE; and (b) a first member of a binding pair, that is linked to
XX the FCE chain; (i) directly by bonds utilizing the N-terminal amino
XX groups, the C-terminal carboxy groups or side-chain functionalities;
XX (ii) via a bifunctional linkage moiety linking the groups or
XX functionalities; or (iii) by the first member binding to the second
XX member of the binding pair, where the second member of the binding pair
XX is covalently attached to the first polypeptide chain. The FCE can be
XX used in a method for producing fibrin. Fibrin is useful as a sealant in
XX surgery to, e.g. reduce bleeding by sealing blood vessels, and tissues

```

```

CC that have been dissected either in surgery or through wounding. The
CC fusion protein allows for the removal of the fibrinogen converting
CC enzyme from the fibrin sealant preparation via the binding of
CC streptavidin to a biotin solid support. The present sequence encodes
CC chicken (Gallus gallus) avidin as given in the present invention.
XX
SQ Sequence 459 BP; 118 A; 131 C; 120 G; 90 T; 0 other;

Query Match      80.1%; Score 389.4; DB 20; Length 459;
Best Local Similarity 98.5%; Pred. No. 5.1e-113;
Matches 393; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 88 ccagagatccctgcagaaagtgcctgctgactgggaaatggaccaacgactcgggctcc 147
Db 61 cccggcctctcgcagaaagtgcctgctgactgggaaatggaccaacgactcgggctcc 120
QY 148 aacatgaccatcgggcgtgtgaacagcagaggtgaattccacagggcacctacacacagcc 207
Db 121 aacatgaccatcgggcgtgtgaacagcagaggtgaattccacagggcacctacacacagcc 180
QY 208 gtaacagccacatcaaatgagatcaaaagagtcaccattgcatggagacacaaacaccatc 267
Db 181 gtaacagccacatcaaatgagatcaaaagagtcaccattgcatggagacacaaacaccatc 240
QY 268 aacaagagagaccagcccaccttggcttcacagtcacaaattggaagttttcagagtcacc 327
Db 241 aacaagagagaccagcccaccttggcttcacagtcacaaattggaagttttcagagtcacc 300
QY 328 actgtcttcacggccagtgcttcacagcaggaatgggaagaggtcctgaagacacatg 387
Db 301 actgtcttcacggccagtgcttcacagcaggaatgggaagaggtcctgaagacacatg 360
QY 388 tggctgctcggtccaagtgttaatgacattggtgatgactggaaggtaccagggctcgcc 447
Db 361 tggctgctcggtccaagtgttaatgacattggtgatgactggaaggtaccagggctcgcc 420
QY 448 atcaaatcttcaactgcctgcgcacacagagaggtga 486
Db 421 atcaaatcttcaactgcctgcgcacacagagaggtga 459

RESULT 4
AAT73194
ID AAT73194 standard; DNA; 604 BP.
XX
XX AAT73194;
XX
XX 27-APR-1998 (first entry)
XX
XX Wild-type avidin DNA.
XX
XX Streptavidin; biotin; anti-interference reagent; detection; mutein;
XX avidin; non-specific binding; ss.
XX
XX Unidentified.
XX
XX Key      Location/Qualifiers
XX CDS      44..502
XX          /*tag= a
XX          /*product= avidin
XX          /*tag= b
XX          /*tag= c
XX          /*product= avidin
XX
XX DE19637718-A1.
XX
XX 02-OCT-1997.
XX
XX 16-SEP-1996; 96DE-1037718.
XX

```



```

Db 2126 gggccagtgcttcacagacgaatgggaagggtccctgaagaccatgtggtgtgcg 2185
      |||
Qy 399 gtcaagtgttaatgacattggtgatgactggaagctaccagggtcggtcgaacacattt 458
      |||
Db 2186 gtcaagtgttaatgacattggtgatgactggaagctaccagggtcggtcgaacacattt 2245
      |||
Qy 459 cactgcctgcgcacacagaagagtgga 486
      |||
Db 2246 cactgcctgcgcacacagaagagtgga 2273
      |||

RESULT 6
AAAT68502
ID AAT68502 standard; DNA; 484 BP.
AC AAT68502;
XX
XX
XX 09-FEB-1998 (first entry)
XX
XX Barley alpha amylase signal sequence and avidin chimeric gene.
XX
XX Avidin; chicken; maize; transgenic plant; plant expression;
KW barley alpha amylase signal sequence; promoter; chimeric gene; ss.
XX
XX Synthetic.
OS Chimeric - Hordeum vulgare.
OS Chimeric - Gallus domesticus.
XX
XX WO9717455-A2.
PN
XX
XX 15-MAY-1997.
XX
XX 31-OCT-1996; 96WO-US17091.
XX
XX 06-NOV-1995; 95US-0554586.
XX
XX (HOWA/) HOWARD J.
XX
XX Baszczynski C, Hood E, Howard J, Maddock S, Meyer TE;
PI Register JC, Witcher D;
XX
XX WPI; 1997-281044/25.
XX
XX Production of avidin in plants - using a heterologous nucleotide
PT sequence encoding avidin operably linked to a promoter to allow
PT expression in plants
XX
XX Example 1; Page 22-23; 28pp; English.
XX
XX The present sequence represents a novel DNA molecule which comprises:
CC (a) a barley alpha amylase promoter; and (b) an optimized DNA sequence
CC derived from the chicken avidin using a codon bias table for maize,
CC which is operably linked to the promoter. A method has also been
CC developed for producing avidin in commercial quantities. The method
CC involves: (a) providing biomass (preferably seeds) from plants, of
CC which at least certain plants contain a DNA molecule comprised of a
CC heterologous nucleotide sequence coding for avidin, where the nucleotide
CC sequence is operably linked to a promoter to effect expression of
CC avidin by certain plants (e.g. the present sequence); and (b) extracting
CC avidin from the biomass. The method can be used for the production of
CC high levels of avidin in native conformation and active form. The
CC constructs can also provide for high level expression of other genes.
XX
XX Sequence 484 BP; 108 A; 182 C; 116 G; 78 T; 0 other;

Query Match 57.6%; Score 279.8; DB 18; Length 484;
Best Local Similarity 82.7%; Pred. No. 2.6e-78;
Matches 320; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 100 gccagaaagtgtcgtgactggaatgaccacgatctgggtcccaatgaccatc 159
      |||

```

```

Db 84 gccaggaagtgtctccctcaccgcgcaagtggaccaatgacctggtcccaatgaccatc 143
Qy 160 gggcctgtgaacacagacagagggtgaattcacaggcacctacatcacacgcgtaaacacaca 219
      |||
Db 144 ggcgccgtgaactccaggggcgagttccacgcgcacctacatcacccgcgtgaccgccacc 203
Qy 220 tcaaatgagatcaaaagatgcaccattgtcatgggacacacacacacacacacacacacacac 279
      |||
Db 204 tccaacgagatcaaggaggtccccctccacggtaccacagaacacacacacacacacacacac 263
Qy 280 cagccacaccttggcttcaccgtcaattgggaagttttcagagtcaccacctgtcttcacg 339
      |||
Db 264 cagccacaccttgcggttcacccgtgaactggaagtcttcgcgagtcaccacgcgtgttcacc 323
Qy 340 ggcagtgcttcacatagacaggaatgggaaggaggtccctggaagaccatgtggtgtcggg 399
      |||
Db 324 ggcagtgcttcacatagacaggaatgggaaggaggtccctggaagaccatgtggtgtcggg 383
Qy 400 tcaagtgttaatgacattggtgatgactggaagctaccagggtcggtcgcacacacacacac 459
      |||
Db 384 agtccctggaatgacatcggcgcgactgggaaggcccccgcggtgggcatcaacatcttc 443
Qy 460 actgcctgcgcacacagaaggagtgga 486
      |||
Db 444 accgcctccgacccagaggagtgga 470
      |||

RESULT 7
AAZ49854
ID AAZ49854 standard; cDNA; 324 BP.
XX
XX AAZ49854;
XX
XX 25-APR-2000 (first entry)
XX
XX Potato proteinase inhibitor-I cDNA.
XX
XX Potato proteinase inhibitor-I; PPI-I; avidin; worm; insect;
KW plant-noxious protein; pest resistance; moth; insect; weevil;
KW grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;
KW insecticidal; ss.
XX
XX Solanum tuberosum.
XX
XX Key Location/Qualifiers
FH 1..324
FT CDS /*tag= a
FT /*product= "Potato proteinase inhibitor-I"
FT sig_peptide 1..69 /*tag= b
FT mat_peptide 70..321 /*tag= c
FT misc_feature 82..103 /*tag= d
FT /*note= "Region used to design mutagenic primer"
XX
XX WO200004049-A1.
PN
XX 27-JAN-2000.
XX
XX 15-JUL-1999; 99WO-NZ00110.
XX
XX 15-JUL-1998; 98NZ-0331002.
XX
XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX
XX Christeller JT, Sutherland PW, Murray C, Markwick NP, Philip BA;
PI Malone LA, Burgess EPJ;
XX
XX WPI; 2000-171244/15.
XX
XX New chimeric polypeptide and composition comprising the polypeptide
PT useful for conferring pest resistance on plants -
PT

```

Example 2; Fig 1; lllpp; English.

The present sequence encodes potato proteinase inhibitor-I (PPI-I). This is used in the preparation of a binary vector designed to express a chimeric polypeptide comprising Avidin mature peptide, a plant-toxicous protein, fused to the PPI-I signal peptide. The binary vector is targeted to the vacuole by PPI-I signal sequence. Transformation of plant genome with the vector can produce pest resistance in plants, a plant derived products and stored harvest material. Pests that can be controlled include, cotton bollworm, tropical army-worm, European corn-borer or red mite, tobacco horn worm, loopers, rice stem borer, porina, cutworms, diamondback moth, potato tuber moth, codling moth, Indian meal moth, gypsy moth, Argentine stem weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat weevils, mealworms, flour beetles, black field cricket, locusts, sawflies, Western flower thrips, Hessian flies or two-spotted mite.

Sequence 324 BP; 101 A; 53 C; 68 G; 102 T; 0 other;

[illegible]

RESULT	8	
AAS85309		
ID	AAS85309	standard; cDNA; 1700 bp.
XX	XX	
AC	AAS85309;	
XX	XX	
DT	13-FEB-2002	(first entry)
XX	XX	
DE	DNA	encoding novel human diagnostic protein #21113.
XX	XX	
KW	Human;	chromosome mapping; gene mapping; gene therapy; forensic;
KW	food	supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	XX	
OS	Homo sapiens.	

polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

[illegible]

```

RESULT          9
ABL08242
ID  ABL08242 standard; cdNA; 2781 BP.
XX
AC  ABL08242;
XX
DT  26-MAR-2002 (first entry)
XX
DE  Drosophila melanogaster expressed polynucleotide SEQ ID NO 19208.
XX
KW  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical; gene; ss.
XX
OS  Drosophila melanogaster.
XX
WC  200171042-A2.
XX
PD  27-SEP-2001.
XX
PF  23-MAR-2001; 2001WO-US09231.
XX
PR  23-MAR-2000; 2000US-191637P.
XX
PR  11-JUL-2000; 2000US-0614150.
XX
(PEKE ) PE CORP NY.
XX
Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI: 2001-656860/75.
XX
P-PSDB; ABB64139.
XX
NEW Isolated nucleic acid detection reagent for detecting 1000 or more

```

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1; SEQ ID NO 19208; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO.
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2781 BP; 760 A; 590 C; 696 G; 735 T; 0 other;

Query Match 7.2%; Score 34.8; DB 23; Length 2781;
 Best Local Similarity 50.8%; Pred. No. 2.1;
 Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 93 gatccctgcagaaagtgcctgactgctggaatgaccaaogactctgggtccaacat 152

Db 124 gctgtctcgagcagtgggcaatgatccggcgatgacgcgttactggtcttgag 183

QY 153 gaccatcggtgtgaacagcagagtgatccaagggcaccctacatcacagccgtaac 212

Db 184 aggaatcgcaagtgtgaacagagcagtcacaaaacttctatcaaatccgcaatgacaag 243

QY 213 agccacatcaaatgagatcaaaagatgcaccattgcatgggacacaa 258

pB 244 accgagagtattatcatatcacacggagaccattgttcggygacacaa 289

RESULT 10

ABL08255

ID ABL08255 standard; cDNA; 4512 BP.

XX ABL08255;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 19247.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB64152.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1; SEQ ID NO 19247; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO.
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4512 BP; 1395 A; 1027 C; 1192 G; 898 T; 0 other;

Query Match 7.2%; Score 34.8; DB 23; Length 4512;
 Best Local Similarity 50.6%; Pred. No. 2.7;
 Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 93 gatccctgcagaaagtgcctgactgctggaatgaccaaogactctgggtccaacat 152

Db 3829 gctgtctcgagcagtgggcaatgatccggcgatgacgcgttactggtcttgag 3888

QY 153 gaccatcggtgtgaacagcagagtgatccaagggcaccctacatcacagccgtaac 212

Db 3889 aggaatcgcaagtgtgaacagagcagtcacaaaacttctatcaaatccgcaatgacaag 3948

QY 213 agccacatcaaatgagatcaaaagatgcaccattgcatgggacacaa 258

Db 3949 accgagagtattatcatatcacacggagaccattgttcggygacacaa 3994

RESULT 11

ABL08254

ID ABL08254 standard; cDNA; 6585 BP.

XX ABL08254;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 19244.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB64151.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

PS Claim 1; SEQ ID NO 19244; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and

SQ sequence 2543 BP; 740 A; 477 C; 557 G; 769 T; 0 other;

	Query Match	6.9%	Score 33.4	DB 21	Length 2543
	Best Local Similarity	57.0%	Prod. No. 5.7		
	Matches	61	Conservative	0	Mismatches 46
					Indels 0
					Gaps 0
QY	202	acagcgttaacagcgcacatcaaatgaagatcaaaaagagtcacacattgcatgggagacacaaaac	261		
Db	2334	aaacagcaacagcaatcatcagagacaaagggctcactgacgcgacacacacacaac	2393		
QY	262	accatcaacaagaggaccagccacctttgtggttcaccgctcaattg	308		
Db	2394	aacacacaacgctctcatctcaattctattcttctcaacaattg	2440		

Search completed: September 20, 2002, 16:25:08
Job time: 3747 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 14:48:21 ; Search time 1842.77 Seconds
(without alignments)
5519.027 Million cell updates/sec

Title: US-09-743-690-6

Perfect score: 486

Sequence: 1 atggagtcacaaagtttgctca.....tgcgcacacagaaggagtgga 486

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	DB ID	Description

ALIGNMENTS

RESULT 1

GGAVIR	GGAVIR	604 bp	mRNA	linear	VRT 21-MAR-1995
LOCUS	Chicken mRNA for avidin.				
DEFINITION	X05343				
ACCESSION	X05343.1	GI:63071			
VERSION	avidin.				
KEYWORDS	chicken.				
SOURCE	Gallus gallus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
REFERENCE	1 (bases 1 to 604)				
AUTHORS	Gope,M.L., Keinanen,R.A., Kristo,P.A., Conneely,O.M., Beattie,W.G., Zarucki-Schulz,T., O'Malley,B.W. and Kulomaa,M.S.				
TITLE	Molecular cloning of the chicken avidin cDNA				
JOURNAL	Nucleic Acids Res. 15 (8), 3595-3606 (1987)				
MEDLINE	87203384				
COMMENT	Data kindly reviewed (18-SEP-1987) by Kulomaa M.S.				
FEATURES	Location/Qualifiers				
SOURCE	1..604				
	/organism="Gallus gallus"				
	/db_xref="taxon:9031"				

X05343 Chicken mRNA
A93651 Sequence 3
AR177253 Sequence
AX018983 Sequence
L27818 Gallus gall
AJ311647 Gallus ga
Z21536 G.gallus av
Z21535 G.gallus av
Z21612 G.gallus av
Z22883 G.gallus av
Z22882 G.gallus av
Z27063 Gallus gall
AJ237659 Gallus ga
Z21611 G.gallus av
AJ237658 Gallus ga
Z21554 G.gallus av
AJ311648 Gallus ga
L06137 Solanum tub
L06606 Solanum tub
X67675 S.tuberosum
X67950 S.tuberosum
L05985 Solanum tub
K03290 Tomato leaf
J05094 L.peruvianu
U30861 Solanum tub
M17108 Potato woun
M13938 Tomato (L.e
Z14027 S.tuberosum
Z12611 S.tuberosum
AC094052 Rattus no
M59427 L.peruvianu
Z12819 S.tuberosum
Z12623 N.cabacum p
X67076 N.cabacum T
AC013353 Trypanoso
AL662911 Mus muscu
AC091912 Homo sapi
AC024186 Homo sapi
AC106776 Homo sapi
AC104042 Homo sapi
AC079073 Homo sapi
Z12619 N.cabacum P
X67075 N.cabacum T
AF429315 Homo sapi
X78997 R.norvegicu


```
Query Match      80.1%; Score 389.4; DB 6; Length 604;
Best Local Similarity 98.5%; Pred. No. 1e-109;
Matches 393; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 88 ccagagatccctccagaaaagtctgcgtgactgggaatggaccaacgatctggggtcc 147
    |||
Db 104 CCGGGCTCTCTGCCAGAAAGTCTCGCTGACTGGGAAATGGACCAACGATCTGGGGTCC 163

QY 148 aacataacacatcgggctgtgaacacagagaggtgaattccacaggaacctacatacaagcc 207
    |||
Db 164 AACATACCATTCGGGGCTGTGAACACAGCAGAGTGAATTCACAGSCACCTACATCACAGCC 223

QY 208 gtaacagcacatcaaatagatcaagagtcaccattgcattggagacacacacacacac 267
    |||
Db 224 GFACAGCCACATCAATAGATCAAGAGTCACCACTGCATGGGACACAAAACACCATC 283

QY 268 aacaagaggaccagccaccttggcttcaccgtcaattgggaagttttcagagtcacc 327
    |||
Db 284 AACAGAGGACCCAGCCACCTTTGGCTTCACCGTCAATTTGGAAGTTTTCAGAGTCCACC 343

QY 328 actgtcttcacggccagtgcttcataagacaggaatgggaagaggtcctcgaagaccatg 387
    |||
Db 344 ACTGTCTTCACGGGCCAGTCTTCATAGACAGGAATGGGAAGGAGGTCTCTGAAGACCATG 403

QY 388 tggctgctcgggtcaagtgtaatgacattggtgactggaagctaccaggggtcggc 447
    |||
Db 404 TGCTGCTGGGTCAAGTGTAAATGACATTTGGTGATGACTGGAAAGCTACCAAGGTCCGC 463

QY 448 atcaacatcttactgcctgcgcacacagaaggagtga 486
    |||
Db 464 ATCAACATCTTACTCGCTCGCCGCACACACAGAGGAGTGA 502

RESULT 4
AX018983
LOCUS      AX018983      5177 bp      DNA      linear      PAT 07-SEP-2000
DEFINITION Sequence 1 from Patent WO9942577.
ACCESSION  AX018983
VERSION     AX018983.1 GI:10043078
KEYWORDS   synthetic construct.
SOURCE      synthetic construct.
ORGANISM    artificial sequence.
REFERENCE   1 (bases 1 to 5177)
AUTHORS     Yla-Herttuala,S., Airenne,K., Kulomaa,M., Marjomaki,V. and
            Lehtolainen,P.
TITLE       Biotin-binding receptor molecules
JOURNAL     Patent: WO 9942577-A 1.26-AUG-1999;
            YLA HERTTUALA SEPPO (FI); AIRENNE KARI (FI); KULOMAA MARKKU (FI);
            MARJOMAKI VARPU (FI); LEHTOLAINEN PAULIINA (FI); EUROGENE LIMITED
            (GB)
FEATURES    Location/Qualifiers
             source
               1..5177
               /organism="synthetic construct"
               /db_xref="taxon:32630"
               /note="Designed DNA sequence to encode transmembrane
               protein."
             CDS
               1071..2273
               /note="unnamed protein product"
               /codon_start=1
               /transl_table=11
               /protein_id="CAC07722.1"
               /db_xref="GI:10043079"
               /translation="NAQWDDFPDQEDTDCSTESVKFDRSVTALLPHPKNGPTLQE
               RMKSKTALITLILYLVFVVLPIGIVAAQLLKWEKNTQVSGVSNADISPSPEKGNG
               SDEMRFRAVMERNMESRLQYLSDNAELLDAKNFQNSITTDQRFNDVLFQLNS
               LLSQTEHNIETGDISKSLVGLNTVLDQFSIETLNGVOENAFKQOEEMKLEERI
               YNASAEKSLDEQVYLEQIEKGMKLLNITNDRKDWHSQTLKNTITLLOGARKC
               SLTGKWTNDLGSMTIGAVNSRGFTGTIVATATSNIEKSPLTGQNTINKRTP
               TFGFTVNMKFSSTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVGINIF
               TRLRTRQE"
```

```
BASE COUNT      1359 a      1221 c      1389 g      1208 t
ORIGIN
Query Match      79.5%; Score 386.4; DB 6; Length 5177;
Best Local Similarity 99.7%; Pred. No. 1.4e-108;
Matches 387; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 99 tccagaaagtgcgtgactgggaatggaccaacgatctgggctccaacatgacct 158
    |||
Db 1886 TGCCAGAAAGTCTCGCTGACTGGGAAATGGACCAACGATCTGGGCTCCAACATGACCAT 1945

QY 159 cgggctgtgaacacagaggtgaattcacagggccctacatcacagcgtaaacagccac 218
    |||
Db 1946 CGGGCTGTGAACACACAGAGTGAATTCACAGGCACCTACATCACAGCCGTACACGCCAC 2005

QY 219 atcaaatgagatcaaaagagtcaccattgcattgggacacacacacacacacacagag 278
    |||
Db 2006 ATCAAAATGAGATCAAAAGAGTCACCACTGCATGGGACACAAAACACCATCAACAAGAGGAC 2065

QY 279 ccagccaccttggcttcaccgtcaattgggaagttttcagagtcaccacctgtcttcac 338
    |||
Db 2066 CCAGCCCACTTTGGCTTCACCGTCAATTTGGAAGTTTTCAGAGTCCACCACTGTCTTCAC 2125

QY 339 gggccagtgctctacagacaggaatgggaagaggtcctcgaagaccatgtgctgctcg 398
    |||
Db 2126 GGGCAGTGTCTATAGACAGGAATGGGAAGGAGTCTCTGAAGACCATGTGCTGCTGGC 2185

QY 399 gtcaagtgttaatgacattggtgactggaagctaccaggggtcggcacatcatt 458
    |||
Db 2186 GTCAAGTGTAAATGACATTTGGTGATGACTGGAAAGCTACCAAGGTTCGGCATCAACATCTT 2245

QY 459 cactgcctgcgcacacagaaggagtga 486
    |||
Db 2246 CACTGCCTGCGCACACAGAGGAGTGA 2273

RESULT 5
CHKAVID
LOCUS      CHKAVID
DEFINITION Gallus gallus (clone 44) STS, avidin gene, complete cds, sequence
            tagged site.
ACCESSION  L27818
VERSION     L27818.1 GI:450255
KEYWORDS   STS; avidin; primer; sequence tagged site.
SOURCE      Gallus gallus (strain White Leghorn, sub_species domesticus) female
            adult oviduct DNA.
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (sites)
AUTHORS     Delange,R.J. and Huang,T.-S.
TITLE       Egg white avidin III. Sequence of the 78-residue middle cyanogen
            bromide peptide. Complete amino acid sequence of the protein
            subunit
            J. Biol. Chem. 246, 698-709 (1971)
JOURNAL     71107558
MEDLINE
REFERENCE   2 (sites)
AUTHORS     Gope,M.L., Keinaenen,R.A., Kristo,P.A., Conneely,O.M.,
            Beattie,W.G., Zarucki-Schulz,T., O'Malley,B.W. and Kulomaa,M.S.
TITLE       Molecular cloning of the chicken avidin cDNA
JOURNAL     Nucleic Acids Res. 15, 3595-3606 (1987)
MEDLINE     87203384
REFERENCE   3 (bases 1 to 1119)
AUTHORS     Wallen,M.J., Laukkanen,M.O. and Kulomaa,M.S.
TITLE       Sequence of the chicken egg-white avidin gene
JOURNAL     Unpublished (1994)
MEDLINE
FEATURES    Location/Qualifiers
             source
               1..1119
               /organism="Gallus gallus"
               /strain="White Leghorn"
               /sub_species="domesticus"
```

```
/db_xref="taxon:9031"
/sex="female"
/tissue_type="oviduct"
/dev_stage="adult"
1. .1119
/note="30 cycles (94C 1 min, 55C 1 min, 72C 1.5 min), Taq
pol., 1119 bp"
1. .21
/note="PCR; putative"
join(13. .93,178. .368,814. .934,1022. .1067)
/codon_start=1
/product="avidin"
/protein_id="AAB59733.1"
/db_xref="GI:451080"
/translation="MVHATSPDLLLLSALVAPLSARKCSLTGKWNDLGSNMTI
GAVNSKGFDTGTTTAVTATSNIEKESPLHGTONIKRTQPTFGFTYNNKFSSTTV
FTQCQFDTRNGKEVLKTMWLLRSSVNDIGDWDKATRVGINIFRLRTOKE"
<13. .93
76
/note="corresponds to position 107 in the cDNA; a possible
PCR error in the gene sequence"
/citation=[2]
/replace="g"
94. .177
178. .388
325. .327
/note="the gene and the cDNA encode for glutamine whereas
the protein sequence has a glutamic acid at this position"
/citation=[1]
/replace="gar"
389. .813
814. .934
935. .1021
1022. .>1119
complement(1100. .1119)
/note="PCR; putative"
210 a 382 c 284 g 243 t
BASE COUNT
ORIGIN
Query Match 43.5% Score 211.4; DB 11; Length 1119;
Best Local Similarity 99.5%; Pred. No. 2e-54; Indels 0; Gaps 0;
Matches 212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 107 agtgcctgactgggaatggaccaacgatctgggtccaaatgaccatcgggctg 166
|||||
Db 176 AGTCCTGCTGACTGGCAATGGACCAACGATCTGGGCTCCAACATGACCATCGGGCTG 235
QY 167 tgaacagcagaggtgaattcaacagccctacatcaacagccgtacacgcccacataaatg 226
Db 236 TGAACAGCAGAGGTCAATTACAGCCACCTACATCAGCCGCTAACAGCCACATCAAAATG 295
QY 227 agatcdaadagatcaccatttcattggacacadaacacacatcaacagggagccagccca 286
|||||
Db 296 AGATCAAGAGTACCCTGCTGCTGGACAAAACACATCAACAGACGACCCGCCCA 355
QY 287 cetttgcttcacgcgtcaattggaattttcag 319
|||||
Db 356 CCATTGGCTTCCAGCTCAATTGGAGTTTCAG 388
RESULT 6
LOCUS GGA311647 1224 bp DNA linear VRT 18-MAR-2001
DEFINITION Gallus gallus AVD gene for avidin, exons 1-4.
ACCESSION AJ311647
VERSION AJ311647.1 GI:13397825
KEYWORDS AVD gene; avidin.
SOURCE chicken
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
```

```
REFERENCE 1 (bases 1 to 1224)
AUTHORS Wallen,M.J., Laukkanen,M.O. and Kulomaa,M.S.
TITLE Cloning and sequencing of the chicken egg-white avidin-encoding
JOURNAL gene and its relationship with the avidin-related genes Avrl-Avr5
MEDLINE Gene 161 (2), 203-209 (1995)
REFERENCE 95394357
AUTHORS 2 (bases 1 to 1224)
Ahliroth,M.K., Kola,E.H., Ewald,D., Masabanda,J., Sozanov,A.,
Pries,R. and Kulomaa,M.S.
TITLE Characterization and chromosomal localization of the chicken avidin
gene family
JOURNAL Anim. Genet. 31 (6), 367-375 (2000)
MEDLINE 21070478
REFERENCE 3 (bases 1 to 1224)
AUTHORS Ahliroth,M.K.
TITLE Direct submission
JOURNAL Submitted (09-MAR-2001) Ahlroth M.K., Department of Biological and
Environmental Science, University of Jyväskylä, PO Box 35,
FIN-40351 Jyväskylä, FINLAND
FEATURES
source Location/Qualifiers
1. .1224
/organism="Gallus gallus"
/db_xref="taxon:9031"
/chromosome="2"
26. .33
join(96. .178,263. .473,899. .1019,1107. .1152)
/gene="AVD"
/codon_start=1
/product="avidin"
/protein_id="CAC34569.1"
/db_xref="GI:13397826"
/translation="MVHATSPDLLLLSALVAPLSARKCSLTGKWNDLGSNMTI
GAVNSKGFDTGTTTAVTATSNIEKESPLHGTONIKRTQPTFGFTYNNKFSSTTV
FTQCQFDTRNGKEVLKTMWLLRSSVNDIGDWDKATRVGINIFRLRTOKE"
98. .178
/gene="AVD"
/number=1
98. .1152
/gene="AVD"
join(<98. .178,263. .473,899. .1019,1107. .>1152)
/gene="AVD"
/product="avidin"
98. .169
/gene="AVD"
179. .262
/gene="AVD"
/number=1
263. .473
/gene="AVD"
/number=2
474. .898
/gene="AVD"
/number=2
899. .1019
/gene="AVD"
/number=3
1020. .1106
/gene="AVD"
/number=3
1107. .1152
/gene="AVD"
/number=4
polya_signal 238 a 410 c 318 g 257 t 1 others
BASE COUNT 1215. .1220
ORIGIN
Query Match 41.7%; Score 202.8; DB 5; Length 1224;
Best Local Similarity 94.6%; Pred. No. 9.8e-52;
Matches 210; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 98 ctgcgacgaagtgcctgcctgactgggaatggaccaacgatctgggtccaaatgacca 157
|||||
```

```

Db 252 CTTGCTACAGTCTCGCTGACTGGGAAATGGGACAAACGACCTGGGCTCAACATGACCA 311
QY 158 tcggggctgaaacagagagtggaattcacagggccacctacatcacagccgtaacagcca 217
Db 312 TCGGGCTGTGAACAGCAAGGTGAATTCACAGGCACCTACACCACGCCGTAAACAGCCA 371
QY 218 catcaaatgagatcaagaagtcaccattgcatgggacacaaacacacatcaacaagagga 277
Db 372 CATCAATGAGATCAAGAGTCACCACTGCATGGGACACAAACACCATCAACAAGAGGA 431
QY 278 ccagccacccttggcttcacgcgtcaattggaagttttcag 319
Db 432 CCCAGCCACCTTTGGCTTCACTGTCAATTTGGAAGTTTTCAG 473

RESULT 7
GGAVR3MRN          GGAVR3MRN          277 bp      mRNA      linear      VRT 16-DEC-1997
LOCUS              G.gallus avr3 mRNA for avidin related gene 3, partial.
DEFINITION        G.gallus avr3 mRNA for avidin related gene 3, partial.
ACCESSION         Z21536
VERSION           Z21536.1 GI:65429
KEYWORDS          avidin; avr3 gene.
SOURCE            chicken.
ORGANISM          Gallus gallus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
                  Phasianinae; Gallus.
REFERENCE         1 (bases 1 to 277)
AUTHORS           Wallen, M.J.
TITLE             Direct Submission
JOURNAL           Submitted (01-FEB-1993) Wallen M. J., University of Tampere,
                  Department of Biomedical Sciences, Tampere, Finland, SF-33101
REFERENCE         2 (bases 1 to 277)
AUTHORS           Keinänen, R.A., Wallen, M.J., Kristo, P.A., Laukkanen, M.O.,
                  Toimela, T.A., Helenius, M.A. and Kulomaa, M.S.
TITLE             Molecular cloning and nucleotide sequence of chicken avidin related
                  genes 1-5
JOURNAL           Eur. J. Biochem. In press
REFERENCE         3 (bases 1 to 277)
AUTHORS           Kunnas, T.A., Wallen, M.J. and Kulomaa, M.S.
TITLE             Induction of chicken avidin and related mRNAs after bacterial
                  infection
JOURNAL           Biochim. Biophys. Acta 1216 (3), 441-445 (1993)
MEDLINE          94092737
FEATURES          Location/Qualifiers
                  source
                    1..277
                    /organism="Gallus gallus"
                    /strain="White Leghorn"
                    /db_xref="taxon:9031"
                    /sex="Female"
                    /dev_stage="Immature, estradiol treated"
                    /tissue_type="intestine"
                    /clone="pCAVR3-fl"
                    1..277
                    /partial
                    /gene="avr3"
                    /standard_name="Avidin related gene 3 mRNA"
                    /citation=[2]
                    1..277
                    /gene="avr3"
                    /gene="avr3"
                    /citation=[3]
                    /replace="t"

BASE COUNT      74 a 73 c 76 g 54 t
ORIGIN

Query Match      33.9%; Score 164.8; DB 5; Length 277;
Best Local Similarity 87.0%; Pred. No. 4.2e-40;
Matches 181; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 279 ccagccacccttggcttcacgcgtcaattggaagttttcagaggtccaccactgttttcac 338

```

```

Db 38  CCACCCACCTTTGGCTTCACTGCTCCATTGGAACCTTTTCAGAGTCCACAGTGTCTTTGT 97
QY 339 gggccagtgcttcatacagagaatgggaagaggtcctgaagaccatgttgctactcgc 398
Db 98  GGGCCAGTGTCTTATAGACAGGCGGAAAGAGGTCTCTGAAGACCAATGGCTGCAACG 157
QY 399 gtcaagtgttaatacattgctgactgctgaaagctaccagggtcggcatacaacatctt 458
Db 158 GTTAGCAGTTGATGACATTAGTACTGATGCTGGAAGCTACCAGGTCGGCTACACAACCT 217
QY 459 cactcgctgcgcacacagaagagtga 486
Db 218 CACTCGCCAGGCACAGTGGAGGAGTGA 245

RESULT 8
GGAVR2MRN          GGAVR2MRN          277 bp      mRNA      linear      VRT 03-FEB-1994
LOCUS              G.gallus avr2 mRNA for avidin related gene 2, partial.
DEFINITION        G.gallus avr2 mRNA for avidin related gene 2, partial.
ACCESSION         Z21535
VERSION           Z21535.1 GI:65428
KEYWORDS          avidin; avr2 gene.
SOURCE            chicken.
ORGANISM          Gallus gallus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
                  Phasianinae; Gallus.
REFERENCE         1 (bases 1 to 277)
AUTHORS           Kunnas, T.A., Wallen, M.J. and Kulomaa, M.S.
TITLE             Induction of chicken avidin and related mRNAs after bacterial
                  infection
JOURNAL           Biochim. Biophys. Acta 1216 (3), 441-445 (1993)
MEDLINE          94092737
FEATURES          Location/Qualifiers
                  source
                    1..277
                    /organism="Gallus gallus"
                    /strain="White Leghorn"
                    /db_xref="taxon:9031"
                    /sex="Female"
                    /dev_stage="Immature, estradiol treated"
                    /tissue_type="oviduct"
                    /clone="pCAVR2-fl"
                    1..277
                    /partial
                    /gene="avr2"
                    /standard_name="Avidin related gene 2 mRNA"
                    /citation=[1]
                    1..277
                    /gene="avr2"

BASE COUNT      75 a 72 c 75 g 55 t
ORIGIN

Query Match      32.1%; Score 156; DB 5; Length 277;
Best Local Similarity 81.8%; Pred. No. 2.3e-37;
Matches 180; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 267 caacaagagagccagccacccttggcttcacgcgtcaattggagttttcagagttccac 326
Db 26  CAAAAGAGCCAGCCAGCCCACTTTGGCTTCACTGTCCATTTGGAACTTTTCAGAGTCCAC 85

```



```

MEDLINE      88260103
REMARK       (sites)
REFERENCE    2 (bases 1 to 1334)
AUTHORS      Keinanen,R.A., Wallen,M.J., Kristo,P.A., Laukkanen,M.O.,
              Toimela,T.A., Helenius,M.A. and Kulomaa,M.S.
TITLE        Molecular cloning and nucleotide sequence of chicken avidin-related
              genes 1-5
JOURNAL      Eur. J. Biochem. 220 (2), 615-621 (1994)
MEDLINE      94170814
REFERENCE    3 (bases 1 to 1334)
AUTHORS      Wallen,M.J.
TITLE        Direct Submission
JOURNAL      Submitted (04-JUN-1993) Wallen M. J., University of Tampere,
              Department of Biomedical Sciences, Tampere, Finland, SF-33101
FEATURES     Location/Qualifiers
              source
                1..1334
                /organism="Gallus gallus"
                /strain="White Leghorn"
                /db_xref="taxon:9031"
                /clone="pgAV2H48"
                /sex="Female"
                /tissue_type="oviduct"
                /clone_lib="genomic library (EMBL4) of Bert W. O'Malley"
                /dev_stage="adult"
                join(<177..294,379..583,1009..1129,1217..>1334)
                /gene="avr4"
                /standard_name="Avidin related gene 4"
                /note="putative exons 1-4"
                /citation=[2]
                join(177..294,379..583,1009..1129,1217..1334)
                /gene="avr4"
BASE COUNT   259 a 440 c 344 g 291 t
ORIGIN
-
Query Match      26.8%; Score 130.2; DB 5; Length 1334;
Best Local Similarity 79.3%; Pred. No. 3.4e-29;
Matches 169; Conservative 0; Mismatches 38; Indels 6; Gaps 1;

QY 107 agtgcgtcgtactgggaatggaccacagatctgggtcccaacatgacctcggggctg 166
Db 377 AGTGCTCGCTGACTGGGAATGGACCAACCTGGGCTCCATCAGCATGACCATCAGGCGTG 436

QY 167 tgaacagcagaggtgaattcacaggcacctacatcacgccgtaacagccacatcaaatg 226
Db 437 TGAACAGCAGAGGTGAATTTCACAGGACCTTACCTCAGCGTGTAGCGGATAATCCAGGAA 496

QY 227 agatcaaaagagtcaccattgcattgggacacaaacaccatcaacagagagccagccca 286
Db 497 ACATCAGCTATCACCACCTGCTTGGGATCCACA-----CAAAAGAGCCAGCCGCCCA 550

QY 287 ccttggttcaccgtcaattggaattttcag 319
Db 551 CCTTTGGCTTCACTGTCCATTTGGAACCTTTTCAG 583

RESULT 11
GGAVR5X14
LOCUS          G.gallus avr5 gene, exons 1-4.      1334 bp      DNA      linear      VRT 18-MAR-2001
DEFINITION    222882
ACCESSION     222882
VERSION       222882.1 GI:311812
KEYWORDS      avdin; avr5 gene.
SOURCE        chicken.
ORGANISM      Gallus gallus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus.
REFERENCE     1 (bases 1 to 1334)
AUTHORS      Keinanen,R.A., Laukkanen,M.L. and Kulomaa,M.S.
TITLE        Molecular cloning of three structurally related genes for chicken
              avdin
JOURNAL      J. Steroid Biochem. 30 (1-6), 17-21 (1988)

```

```

MEDLINE      88260103
REMARK       (sites)
REFERENCE    2 (bases 1 to 1334)
AUTHORS      Keinanen,R.A., Wallen,M.J., Kristo,P.A., Laukkanen,M.O.,
              Toimela,T.A., Helenius,M.A. and Kulomaa,M.S.
TITLE        Molecular cloning and nucleotide sequence of chicken avidin-related
              genes 1-5
JOURNAL      Eur. J. Biochem. 220 (2), 615-621 (1994)
MEDLINE      94170814
REFERENCE    3 (bases 1 to 1334)
AUTHORS      Wallen,M.J.
TITLE        Direct Submission
JOURNAL      Submitted (04-JUN-1993) Wallen M. J., University of Tampere,
              Department of Biomedical Sciences, Tampere, Finland, SF-33101
FEATURES     Location/Qualifiers
              source
                1..1334
                /organism="Gallus gallus"
                /strain="White Leghorn"
                /db_xref="taxon:9031"
                /clone="pgAV2H52"
                /sex="Female"
                /tissue_type="oviduct"
                /clone_lib="genomic library (EMBL4) of Bert W. O'Malley"
                /dev_stage="adult"
                join(<177..294,379..583,1009..1129,1217..>1334)
                /gene="avr5"
                /standard_name="Avidin related gene 5"
                /note="putative exons 1-4"
                /citation=[2]
                join(177..294,379..583,1009..1129,1217..1334)
                /gene="avr5"
BASE COUNT   260 a 440 c 343 g 291 t
ORIGIN
-
Query Match      26.8%; Score 130.2; DB 5; Length 1334;
Best Local Similarity 79.3%; Pred. No. 3.4e-29;
Matches 169; Conservative 0; Mismatches 38; Indels 6; Gaps 1;

QY 107 agtgcgtcgtactgggaatggaccacagatctgggtcccaacatgacctcggggctg 166
Db 377 AGTGCTCGCTGACTGGGAATGGACCAACCTGGGCTCCATCAGCATGACCATCAGGCGTG 436

QY 167 tgaacagcagaggtgaattcacaggcacctacatcacgccgtaacagccacatcaaatg 226
Db 437 TGAACAGCAGAGGTGAATTTCACAGGACCTTACCTCAGCGTGTAGCGGATAATCCAGGAA 496

QY 227 agatcaaaagagtcaccattgcattgggacacaaacaccatcaacagagagccagccca 286
Db 497 ACATCAGCTATCACCACCTGCTTGGGATCCACA-----CAAAAGAGCCAGCCGCCCA 550

QY 287 ccttggttcaccgtcaattggaattttcag 319
Db 551 CCTTTGGCTTCACTGTCCATTTGGAACCTTTTCAG 583

RESULT 12
GGAVR1
LOCUS          Gallus gallus mRNA for avidin related protein 1, partial. 170 bp      mRNA      linear      VRT 18-MAR-2001
DEFINITION    297063
ACCESSION     297063
VERSION       297063.1 GI:2226407
KEYWORDS      avdin related protein; avr1 gene.
SOURCE        chicken.
ORGANISM      Gallus gallus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus.
REFERENCE     1 (bases 1 to 170)
AUTHORS      Keinanen,R.A., Wallen,M.J., Kristo,P.A., Laukkanen,M.O.,
              Toimela,T.A., Helenius,M.A. and Kulomaa,M.S.
TITLE        Molecular cloning and nucleotide sequence of chicken avidin-related
              genes 1-5

```


D	b	167	CTTCGTACAGTGC	TGCGTGACTGGGAATGGGACAACAACCTGGGTCCCAACATGACCA	226
Oy	158	tccgggctgtgaacagcagcgadgtgaatttcacagcgacctatcatcacagcgcaaacagcca	217		
D	b	227	TCGGAGCTGTGAACGACAATGCGCAGITCAATGCGACCTACATCACAGCTGTACGAGATA	286	
Oy	218	catacaaatgaagatacaagagtcaccatttgatgcgatgggacacaacaaacaccatcaacaagagga	277		
D	b	287	ATCCAGGAAAACATCAAGCATACCACTGCTTGGGATCCCAACA-----CAAAGAAGCCT	340	
Oy	278	ccagccacacttggcttcccggtcaattggaagttttcag	319		
D	b	341	GCCAGCCACCCTTTGGCTTCACTGCCATTGGAACCTTTTCAG	382	
RESULT 14					
GGAVRIGEN					
LOCUS					
DEFINITION G.gallus avr1 gene.					
ACCESSION Z21611.1 Z21623					
VERSION Z21611.1 GI:65431					
KEYWORDS avridin; avr1 gene.					
SOURCE chicken.					
ORGANISM Gallus gallus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;					
Phasianinae; Gallus.					
1 (bases 1 to 1335)					
Keinanen,R.A., Laukkanen,M.L. and Kulomaa,M.S.					
Molecular cloning of three structurally related genes for chicken					
avridin					
J. Steroid Biochem. 30 (1-6), 17-21 (1988)					
JOURNAL	88260103				
MEDLINE	94170814				
REMARK					
REFERENCE	2 (bases 1 to 1335)				
AUTHORS	Keinanen,R.A., Wallen,M.J., Kristo,P.A., Laukkanen,M.O., Toimela,T.A., Helenius,M.A. and Kulomaa,M.S.				
TITLE	Molecular cloning and nucleotide sequence of chicken avridin-related genes 1-5				
JOURNAL	Eur. J. Biochem. 220 (2), 615-621 (1994)				
MEDLINE	94170814				
REFERENCE	3 (bases 1 to 1335)				
AUTHORS	Wallen,M.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-1993) Wallen M. J., University of Tampere, Department of Biomedical Sciences, Tampere, Finland, SF-33101				
FEATURES	Location/Qualifiers				
source	1. .1335				
	/organism="Gallus gallus"				
	/strain="White Leghorn"				
	/db_xref="taxon:9031"				
	/clone="pgAV1.8"				
	/sex="Female"				
	/tissue_type="oviduct"				
	/clone_lib="genomic library (EMBL4) of Bert W. O'Malley"				
	/dev_stage="adult"				
mRNA	join(<177..294,379..583,1010..1130,1218..>.1335)				
	/gene="avr1"				
	/standard_name="Avidin related gene 1"				
	/notes="putative exons 1-4"				
	/citation=[1]				
	<177..294				
exon	/gene="avr1"				
	/number=1				
gene	join(177..294,379..583,1010..1130,1218..>.1335)				
	/gene="avr1"				
exon	379..583				
	/gene="avr1"				
	/number=2				
exon	1010..1130				
	/gene="avr1"				
	/number=3				
exon	1218..>.1335				

```

/product="avidin-related protein 6 (AVR6)"
/protein_id="CAB39893.1"
/db_xref="GI:4539028"
/db_xref="SWISS-PROT:P56735"
/translation="MVHATSPDLLLLLSLALVAPGLSARKCSLTGEWNNLGSIMTI
GAVNDGEFNGTAVTAVADPCNTKIGDLICIOHKRACDPFCPTVHNFSESTSVFV
GOCFVDRSGKEYLTKRWLQLAYDDISDDPHKATRYGYNNTFRQTVBE"
<13..84

```

sig_peptide

mat_peptide

intron

exon

intron

exon

intron

exon

intron

exon

polyA_signal

BASE COUNT

ORIGIN

201 a 388 c 293 g 251 t

```

Query Match      24.0%; Score 116.8; DB 5; Length 1133;
Best Local Similarity 73.9%; Pred. No. 4.7e-25;
Matches 164; Conservative 0; Mismatches 52; Indels 6; Gaps 1;

Qy 98 CTGCCAGAAAGTCTCTGCTGACTGGGAAATGGACCAAGATCTGGGTCCTCAACATGACCA 157
   ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 167 CTTCTCACAGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 226
   ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 158 TCGGGGCTGTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 217
   ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 227 TCGGGGCTGTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 286
   ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 218 CATCAATGAGATCAAGAGTCAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 277
   ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 287 ATCCAGGAACATCAAGCTATCAAGCTATCAAGCTATCAAGCTATCAAGCTATCAAGCTATCA 340
   ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 278 CCCAGCCACCTTGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 319
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 GCCAGCCACCTTGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 382

```

Search completed: September 20, 2002, 16:20:58
Job time: 5557 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 15:41:14 ; Search time 27.21 Seconds
(without alignments)
1023.601 Million cell updates/sec

Title: US-09-743-690-7

Perfect score: 845

Sequence: 1 MESKFAHIIIVFFLLATPFET.....DWKATRVGINIFRLRTQKE 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	670	79.3	152	13 Q98SH4	Q98sh4 gallus gall
2	151	17.9	107	10 Q07459	Q07459 solanum tub
3	151	17.9	107	10 Q43648	Q43648 solanum tub
4	146	17.3	106	10 Q43651	Q43651 solanum tub
5	136	16.1	529	5 Q25058	Q25058 heliocidari
6	129	15.3	406	5 Q25059	Q25059 heliocidari
7	122	14.4	101	10 Q41434	Q41434 solanum tub
8	97	11.5	30	10 Q41490	Q41490 solanum tub
9	95	11.2	1395	8 Q35059	Q35059 marchantia
10	84.5	10.0	400	5 Q9VQY1	Q9vqy1 drosophila
11	84.5	10.0	749	5 Q967D9	Q967d9 drosophila
12	84.5	10.0	903	5 Q967D8	Q967d8 drosophila
13	84.5	10.0	1531	5 Q967D7	Q967d7 drosophila
14	82.5	9.8	1153	10 Q64728	Q64728 arabidopsis
15	82.5	9.8	1218	10 Q9LQK4	Q9lqk4 arabidopsis
16	81	9.6	570	3 Q9HFF8	Q9hff8 saccharomyc

17	80	9.5	1592	10	Q9IGT4	Q9lgt4 oryza sativ
18	78.5	9.3	445	5	Q9W0F6	Q9w0f6 drosophila
19	78.5	9.3	802	5	P91151	P91151 caenorhabdi
20	77.5	9.2	434	11	Q35317	Q35317 mus musculu
21	77.5	9.2	475	11	Q62664	Q62664 rattus norv
22	77.5	9.2	1124	4	Q60280	Q60280 homo sapien
23	77	9.1	401	6	Q08835	Q08835 cercopitheci
24	75.5	8.9	263	3	Q13854	Q13854 schizosacch
25	75.5	8.9	539	2	Q9RK08	Q9rkq8 streptomyce
26	75	8.9	465	10	Q9CAN8	Q9can8 arabidopsis
27	75	8.9	693	17	Q58898	Q58898 pyrococcus
28	75	8.9	1441	9	Q9AZL2	Q9azl2 bacteriophia
29	75	8.9	1441	16	Q9CFH5	Q9cfh5 lactococcus
30	75	8.9	3201	5	Q9W0U2	Q9w0u2 drosophila
31	74.5	8.8	974	5	Q01574	Q01574 caenorhabdi
32	74	8.8	398	2	Q9RF12	Q9rf12 clostridium
33	74	8.8	424	16	Q9I456	Q9i456 pseudomonas
34	74	8.8	527	5	Q27471	Q27471 caenorhabdi
35	74	8.8	684	4	Q9UF31	Q9uf31 homo sapien
36	74	8.8	3542	5	Q9U5M2	Q9u5m2 plasmodium
37	73.5	8.7	253	4	Q96PG0	Q96pg0 homo sapien
38	73.5	8.7	531	10	Q9SZF7	Q9szf7 arabidopsis
39	73.5	8.7	725	16	Q9CFY6	Q9cfy6 lactococcus
40	73.5	8.7	1250	5	Q23409	Q23409 caenorhabdi
41	73	8.6	336	5	Q19962	Q19962 caenorhabdi
42	73	8.6	480	10	Q944R8	Q944r8 arabidopsis
43	73	8.6	632	10	Q9SM30	Q9sm30 taraxacum o
44	72.5	8.6	460	10	Q9LIF8	Q9lif8 arabidopsis
45	72.5	8.6	481	16	Q9RWA3	Q9rwa3 deinococcus

ALIGNMENTS

RESULT 1

Q98SH4 ID Q98SH4 PRELIMINARY; PRT; 152 AA.
AC Q98SH4:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AVIDIN PRECURSOR.
GN AVID.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Ahlroth M.K.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21070478; PubMed=11167523;
RA Ahlroth M.K.; Koia E.H.; Ewald D.; Masabanda J.; Sazanov A.; Fries R.,
RA Kulomaa M.S.;
RT "Characterization and Chromosomal Localization of the Chicken Avidin
Gene Family.";
RN Anim. Genet. 31:367-375(2000).
RP SEQUENCE FROM N.A.
RX MEDLINE=95394357; PubMed=7665080;
RA Wallen M.J.; Laakkonen M.O.; Kulomaa M.S.;
RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene
and its relationship with the avidin-related genes Avrl-Avr5.";
RL Gene 161:205-209(1995).
DR EMBL; AJ311647; CAC34569.1; -.
DR HSSP; P20701; IAVD.
DR InterPro; IPR000088; Avidin.
DR Pfam; PF01382; Avidin; 1.
DR ProDom; PD016055; Avidin; 1.
DR PROSITE; PS00577; AVIDIN; 1.

```

KW Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 152 AA: 16742 MW: 04B4965B070B974B CRC64:

Query Match 79.3%; Score 670; DB 13; Length 152;
Best Local Similarity 95.5%; Pred. No. 3.2e-57;
Matches 126; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 30 PEIPARKSLTGKWTNDLGSNMTTICAVNSRGEFTGTYITAVTATSNKESPLHGTONVI 89
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 21 PGLSARKSLTGKWDNDLGSNMTTICAVNSKGEFTGTYITAVTATSNKESPLHGTONVI 80
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 90 NKRPQTFEGFVNNKFSSTVFTGQCIFDRNGKEVLKTMMLLRSSVNDIGDDWKATRVG 149
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 81 NKRPQTFEGFVNNKFSSTVFTGQCIFDRNGKEVLKTMMLLRSSVNDIGDDWKATRVG 140
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 150 INIFTRLTQKE 161
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 141 INIFTRLTQKE 152
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 2
QY7459
ID Q07459 PRELIMINARY; PRT; 107 AA.
AC Q07459;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PROTEINASE INHIBITOR I.
GN PINI.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ILAM HARDY; TISSUE=Potato TUBER;
RX MEDLINE=95106304; PubMed=7807552;
RA Reuning L.L., Spriggs T.W., Christeller J.T.;
RT "Evolution of the proteinase inhibitor I family and apparent lack of
RL Plant Physiol. 102:1061-1061(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ILAM HARDY;
RX MEDLINE=95106304; PubMed=7807552;
RA Reuning L.L., Spriggs T.W., Christeller J.T.;
RT "Evolution of the proteinase inhibitor I family and apparent lack of
RL J. Mol. Evol. 39:644-654(1994).
DR EMBL; X67950; CAA48136.1; -
DR ENBL; X67675; CAA47907.1; -
DR HSSP; P19873; IHYM.
DR InterPro; IPR000864; Potato_inhibit.
DR Pfam; PF00280; potato_inhibit; 1.
DR PRINTS; PR00292; POTATOINHBT.
DR PRODOM; PD002604; Potato_inhibit; 1.
DR PROSITE; PS00285; POTATO_INHIBITOR; 1.
SQ SEQUENCE 107 AA; 12148 MW; DAB99CF6F1D8811 CRC64;

Query Match 17.9%; Score 151; DB 10; Length 107;
Best Local Similarity 93.8%; Pred. No. 3.2e-07;
Matches 30; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MESKFAHIIVFFLLATPFETLLARKESDGPET 32
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1 MESKFAHIIVFFLLATPFETLLARKESDGPET 32
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 3
QY7459
ID Q07459 PRELIMINARY; PRT; 107 AA.
AC Q07459;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PROTEINASE INHIBITOR I.
GN PINI.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ILAM HARDY; TISSUE=Potato TUBER;
RX MEDLINE=95106304; PubMed=7807552;
RA Reuning L.L., Spriggs T.W., Christeller J.T.;
RT "Evolution of the proteinase inhibitor I family and apparent lack of
RL J. Mol. Evol. 39:644-654(1994).
DR EMBL; X67950; CAA48136.1; -
DR ENBL; X67675; CAA47907.1; -
DR HSSP; P19873; IHYM.
DR InterPro; IPR000864; Potato_inhibit.
DR Pfam; PF00280; potato_inhibit; 1.
DR PRINTS; PR00292; POTATOINHBT.
DR PRODOM; PD002604; Potato_inhibit; 1.
DR PROSITE; PS00285; POTATO_INHIBITOR; 1.
SQ SEQUENCE 107 AA; 12148 MW; DAB99CF6F1D8811 CRC64;

Query Match 17.9%; Score 151; DB 10; Length 107;
Best Local Similarity 93.8%; Pred. No. 3.2e-07;
Matches 30; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MESKFAHIIVFFLLATPFETLLARKESDGPET 32
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1 MESKFAHIIVFFLLATPFETLLARKESDGPET 32
   I : |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 3
QY7459
ID Q07459 PRELIMINARY; PRT; 107 AA.
AC Q07459;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PROTEINASE INHIBITOR I.
GN PINI.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ILAM HARDY; TISSUE=Potato TUBER;
RX MEDLINE=95106304; PubMed=7807552;
RA Reuning L.L., Spriggs T.W., Christeller J.T.;
RT "Evolution of the proteinase inhibitor I family and apparent lack of
RL J. Mol. Evol. 39:644-654(1994).
DR EMBL; X67950; CAA48136.1; -
DR ENBL; X67675; CAA47907.1; -
DR HSSP; P19873; IHYM.
DR InterPro; IPR000864; Potato_inhibit.
DR Pfam; PF00280; potato_inhibit; 1.
DR PRINTS; PR00292; POTATOINHBT.
DR PRODOM; PD002604; Potato_inhibit; 1.
DR PROSITE; PS00285; POTATO_INHIBITOR; 1.
SQ SEQUENCE 107 AA; 12148 MW; DAB99CF6F1D8811 CRC64;

```

Query Match 17.3%; Score 146; DB 10; Length 106;
Best Local Similarity 93.5%; Pred. No. 9.7e-07;
Matches 29; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 ESKFAHIVFFLLATPFELLARKESDGPFI 32
DB 1 ESKFAHIVFFLLATSFELLARKESDGPV 31

RESULT 5
Q25058 PRELIMINARY; PRT; 529 AA.
AC Q25058;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE FIBROPELLIN IA (FRAGMENT).
OS Helicoidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Echinometridae;
OC Helicoidaris.
OX NCBI_TaxID=7634;
RN [1]
RA Bisgrove B.W.;
RT "Evolution of the fibropellin gene family and pattern of fibropellin
gene expression in sea urchin phylogeny.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L33861; AAA29995.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000088; Avidin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_II.
DR InterPro; IPR001438; EGF_1.
DR InterPro; IPR000083; fibronectin_type_1.
DR Pfam; PF01382; Avidin; 1.
DR Pfam; PF00008; EGF; 7.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00012; FNTYPEI.
DR ProDom; PD016055; Avidin; 1.
DR SMART; SM00179; EGF_CA; 9.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_8.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_11.
DR PROSITE; PS01186; EGF_2; 10.
DR PROSITE; PS01187; EGF_CA; 7.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
FT NON_TER 1
SQ SEQUENCE 529 AA; 55543 MW; D4AE958FCF9ACB5A CRC64;

Query Match 16.1%; Score 136; DB 5; Length 529;
Best Local Similarity 34.5%; Pred. No. 6.4e-05;
Matches 41; Conservative 16; Mismatches 58; Indels 4; Gaps 3;
QY 37 CSLTGKWTNDLGSNMTGAVNSRGFTGYITAVTATSNKESPLHGTONTKRTOPT 96
DB 404 CDEGVWYNECDQITI-TKTSTGMILGDMTAVEITAVGAAPTVMVGYAS--NNYDFPS 460
QY 97 FGFTVNWKFSESTVFTGQCFFIDRNGKEVLKTMWLLRSSVNDIGDDWKAPRVGINIFTR 155
DB 461 FGFTVVRDNGMSTTSTAQCHLCDN-EEVLYTTWINTNMVDTQDQIKSNWVGQDKWTR 518
RESULT 6
Q25059 PRELIMINARY; PRT; 406 AA.
ID Q25059;
AC Q25059;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FIBROPELLIN III (FRAGMENT).
OS Helicoidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Echinometridae;
OC Helicoidaris.
OX NCBI_TaxID=7634;
RN [1]
RA Bisgrove B.W.;
RT "Evolution of the fibropellin gene family and pattern of fibropellin
gene expression in sea urchin phylogeny.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L33862; AAA29996.1; -.
DR HSSP; P00740; IEDM.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000088; Avidin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000083; fibronectin_type_1.
DR Pfam; PF01382; Avidin; 1.
DR Pfam; PF00008; EGF; 7.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00012; FNTYPEI.
DR ProDom; PD016055; Avidin; 1.
DR SMART; SM00179; EGF_CA; 7.
DR SMART; SM00001; EGF_like; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_7.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_7.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 5.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
FT NON_TER 1
SQ SEQUENCE 406 AA; 43475 MW; EC89CCE8ABB7E89A CRC64;

Query Match 15.3%; Score 129; DB 5; Length 406;
Best Local Similarity 32.8%; Pred. No. 0.00022;
Matches 41; Conservative 20; Mismatches 48; Indels 16; Gaps 6;
QY 37 CSLTGKWTNDLGSNMTGAVNSRGFTGYITAVTATSNKESPL-HGTQNTI---NK 91
DB 281 CDEGVWYNECDQITI-SKYSTGMMLGDHLT-----FNERLIGYCAPTVWVGYASK 331
QY 92 RTQ-PTFGFTVNWKFSESTVFTGQCFFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVGI 150
DB 332 HYEFPSEFGFTVVRDNGRSTTSMTGCHLCA-GQEVLYTTWIESNMVDTCEDIKANKVGO 390
QY 151 NIFTR 155
DB 391 DKWTR 395
RESULT 7
Q41434 PRELIMINARY; PRT; 101 AA.
ID Q41434
AC Q41434;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE FIBROPELLIN III (FRAGMENT).
OS Helicoidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Echinometridae;
OC Helicoidaris.
OX NCBI_TaxID=7634;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-KESZTHELYI 855 (WHITE LADY); TISSUE=ETIOLATED SHOOT;
 RX MEDLINE=96198594; PubMed=8925888;
 RA Banfalvi Z., Molnar A., Molnar G., Lakatos L., Szabo L.
 RT "Starch synthesis, and tuber storage protein genes are differently
 RL expressed in Solanum tuberosum and in Solanum brevidens";
 DR EMBL: U30861; AAC49603.1; -.
 DR HSPF: F19873; IHXM.
 DR InterPro: IPR000864; Potato_inhibit.
 DR Pfam: PF00292; potato_inhibit.1.
 DR PRINTS: PR00292; POTATOINHBT.
 DR PRODOM: PD002604; Potato_inhibit.1.
 DR PROSITE: PS00285; POTATO_INHIBITOR; UNKNOWN_1.
 FT NON_TER
 SQ SEQUENCE 101 AA; 11396 MW; A27EG6CADE4D1CEA CRC64;

Query Match 14.4%; Score 122; DB 10; Length 101;
 Best Local Similarity 92.3%; Pred. NO. 0.00019;
 Matches 24; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 7 HIIIVFLLATPFETLLARKESDGPEI 32
 DB 1 HIIIVFLLATPFETLLARKESDGPEV 26

RESULT 8
 Q41490
 ID Q41490 PRELIMINARY; PRT; 30 AA.
 AC Q41490;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PROTEINASE INHIBITOR I (FRAGMENT).
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_taxid=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. RUSSET BURBANK;
 RA Lee J.S., Yang Y.S.;
 RT "Nucleotide sequence of 5' flanking region of potato proteinase
 RL inhibitor I gene";
 DR EMBL: J01192; to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z14027; CAA78402.1; -.
 FT NON_TER 30
 SQ SEQUENCE 30 AA; 3439 MW; 6090AB9BDA387D40 CRC64;

Query Match 11.5%; Score 97; DB 10; Length 30;
 Best Local Similarity 76.9%; Pred. NO. 0.011;
 Matches 20; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 MESKFAHIIIVFLLATPFETLLARKE 26
 DB 1 MEAKFAHIIIVFLLATSFETFNQSE 26

RESULT 9
 Q35059
 ID Q35059 PRELIMINARY; PRT; 1395 AA.
 AC Q35059;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE TRNA-GLY PROTEIN.
 GN TRNA-GLY.
 GN Marchantia polymorpha (Liverwort).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
 OX Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;

OC Marchantiales; Marchantiaceae; Marchantia.
 OX NCBI_taxid=3197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92114051; PubMed=17311062;
 RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
 RA Kohchi T., Ogura Y., Kanegae T., Akashi K., Ohyama K.,
 RT "Gene organization deduced from the complete sequence of liverwort
 RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
 RT mitochondrial genome";
 RL J. Mol. Biol. 223:1-7(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92319654; PubMed=1620617;
 RA Takemura M., Oda K., Yamato K., Ohta E., Nakamura Y., Nozato N.,
 RA Akashi K., Ohyama K.;
 RT "Gene clusters for ribosomal proteins in the mitochondrial genome of a
 RT liverwort, Marchantia polymorpha";
 RL Nucleic Acids Res. 20:3199-3205(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92350681; PubMed=1641342;
 RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
 RA Akashi K., Ohyama K.;
 RT "Transfer RNA genes in the mitochondrial genome from a liverwort,
 RT Marchantia polymorpha: the absence of chloroplast-like tRNAs";
 RL Nucleic Acids Res. 20:3773-3777(1992).
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 CC GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY
 CC SUBUNIT (BY SIMILARITY).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
 DR EMBL: M08929; AAC09444.1; -.
 DR InterPro: IPR000410; ATP-synt_ab_N.
 DR InterPro: IPR000194; ATPase_alpha_beta.
 DR InterPro: IPR000442; Intron_maturase2.
 DR InterPro: IPR000477; RVTse.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00006; ATP-synt_ab; 1.
 DR Pfam: PF02874; ATP-synt_ab_N; 1.
 DR Pfam: PF01348; Intron_maturase2; 1.
 DR Pfam: PF00078; rvt; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR ATP synthesis: ATP-binding; CF(1); Hydrogen ion transport; Hydrolase;
 KW Mitochondrion; RNA-directed DNA polymerase.
 SQ SEQUENCE 1395 AA; 156997 MW; 33DD91CF184F9686 CRC64;

Query Match 11.2%; Score 95; DB 8; Length 1395;
 Best Local Similarity 24.8%; Pred. NO. 1.9;
 Matches 41; Conservative 23; Mismatches 71; Indels 30; Gaps 7;

OY 10 VFFLLATPFETLLARKESDG----PEIDARKCSLTGKWTNDIGSNMTIGAVNSRGFTGT 65
 DB 300 VFYLSRLLERAAKMSDQTGAGSLTALPIECATGR---SHIQWYGVGEVAGRYQT 356

OY 66 YITAVTATSNKESPLHGTQNTINKRTQPTGPTVNM-----PFSEST-----TV 111
 DB 357 RSNPKDQSARRIGRADIHGVOKPIKPSDPVNL-VOWNAASSKYDESSGKHPIPKI 415

OY 112 FTGOCFTDRNKEVLKTMWLNR-SSVNDIGDDWKATRVGINIFTR 155

DB 416 HSGSAHLTRQGR-----LLRPSGVVTEGDKPKAYALGTRNVSK 453

RESULT 10
 Q9VOY1
 ID Q9VOY1 PRELIMINARY; PRT; 400 AA.
 AC Q9VOY1;

Search completed: September 17, 2002, 15:43:22
Job time: 128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 15:41:34 ; Search time 11.91 Seconds
(without alignments)
523.413 Million cell updates/sec

Title: US-09-743-690-7

Perfect score: 845

Sequence: 1 MESKFAHIIIVFFLLATPFET.....DWKATRVGINIFRLRTQKE 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	681	80.6	152	1	AVID_CHICK
2	499	59.1	150	1	AVR4_CHICK
3	469	55.5	150	1	AVR3_CHICK
4	467	55.3	150	1	AVR7_CHICK
5	458	54.2	150	1	AVR1_CHICK
6	457	54.1	150	1	AVR6_CHICK
7	444	52.5	150	1	AVR2_CHICK
8	154	18.2	107	1	IC1D_SOLTU
9	143	16.9	107	1	IC1I_SOLTU
10	143	16.9	111	1	IC1I_LYCES
11	142.5	16.9	183	1	SAV1_STRVL
12	142.5	16.9	183	1	SAV_STRAV
13	140	16.6	1064	1	FBP1_STRPU
14	139.5	16.5	183	1	SAV2_STRVL
15	136	16.1	111	1	IC1I_LYCPPE
16	124	14.7	570	1	FBP3_STRPU
17	85	10.1	481	1	UGDH_CAEEL
18	80.5	9.5	701	1	EPG_MICLU
19	77.5	9.2	434	1	PBX3_HUMAN
20	74.5	8.8	1268	1	PGCN_MOUSE
21	74	8.8	872	1	GUXA_CELFI
22	73.5	8.7	924	1	KE2_CANAL
23	72	8.5	331	1	TRPD_BUCSC
24	71.5	8.5	430	1	PBX1_HUMAN
25	71.5	8.5	729	1	PET2_HUMAN
26	70.5	8.3	630	1	MUC1_MOUSE
27	69.5	8.2	542	1	FIBP_ADECT
28	69.5	8.2	713	1	CDG2_PAEWA
29	69.5	8.2	1286	1	RPD_GUITH
30	68.5	8.1	292	1	MURE_BUCAP
31	68.5	8.1	435	1	VGLM_VZVD
32	68.5	8.1	535	1	ICAI_BOVIN
33	68	8.0	332	1	IPAD_SHIDY
34	68	8.0	1953	1	BIGA_SALTY
35	67.5	8.0	489	1	RHAB_ECOLI
36	67.5	8.0	565	1	YHXB_BACSU
37	67.5	8.0	1852	1	CCAS_CYPCA
38	67	7.9	482	1	K2C8_HUMAN
39	67	7.9	1033	1	TIR1_ECOLI
40	66.5	7.9	210	1	RA10_YEAST
41	66.5	7.9	306	1	C181_HUMAN
42	66	7.8	369	1	YMG6_YEAST
43	66	7.8	417	1	PVR_CERAE
44	66	7.8	911	1	CTDI_MOUSE
45	65.5	7.8	393	1	PGLR_PRUPE

ALIGNMENTS

RESULT 1
AVID_CHICK
ID AVID_CHICK STANDARD; PRT; 152 AA.
AC P02701; Q91958; PRT; 152 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Avidin precursor.
GN AVD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87203384; PubMed=3575102;
RA Gope M.L., Keinänen R.A., Kristo P.A., Conneely O.M., Beattie W.G.,
RA Zarucki-Schulz T., O'Malley B.W., Kulomaa M.S.;
RT "Molecular cloning of the chicken avidin cDNA";
RL Nucleic Acids Res. 15:3595-3606(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90355928; PubMed=2143802;
RA Chandra G., Gray J.G.;
RT "Cloning and expression of avidin in Escherichia coli";
RL Meth. Enzymol. 184:70-79(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=WHITE LEGHORN; TISSUE=Oviduct;
MEDLINE=95394357; PubMed=7665080;
RA Wallen M.J., Laukkanen M.O., Kulomaa M.S.;
RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene and its relationship with the avidin-related genes Avrl-Avr5";
RL Gene 161:205-209(1995).
RN [4]
RP SEQUENCE OF 25-152.
RX MEDLINE=71107558; PubMed=5100763;
RA Delange R.J., Huang T.-S.;
RT "Egg white avidin. 3. Sequence of the 78-residue middle cyanogen bromide peptide. Complete amino acid sequence of the protein subunit";
RL J. Biol. Chem. 246:698-709(1971).
RN [5]
RP IMPORTANCE OF TYR IN BIOTIN-BINDING.
RX MEDLINE=90351377; PubMed=2386489;
RA Gitlin G., Bayer E.A., Wilchek M.;
RT "Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine residues are involved in the binding site";
RL Biochem. J. 269:527-530(1990).
RN [6]
RP BIOTIN-BINDING STUDIES.
RX MEDLINE=91378911; PubMed=1898347;
RA Hiller Y., Bayer E.A., Wilchek M.;
RT "Studies on the biotin-binding site of avidin. Minimized fragments that bind biotin";

RL Biochem. J. 278:573-585(1991).
 RN [7]
 RD X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=93281699; PubMed=8506353;
 RA Livnah O., Bayer E.A., Wilchek M., Sussman J.L.;
 RT "Three-dimensional structures of avidin and the avidin-biotin
 complex.";
 RL PROC. Natl. Acad. Sci. U.S.A. 90:5076-5080(1993).
 RN [8]
 RD X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE=93294833; PubMed=8515446;
 RA Pugliese L., Coda A., Malcovati M., Bolognesi M.;
 RT "Three-dimensional structure of the tetragonal crystal form of
 egg-white avidin in its functional complex with biotin at 2.7-A
 resolution.";
 RL J. Mol. Biol. 231:698-710(1993).
 RN [9]
 RD X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=98430987; PubMed=9760187;
 RA Nardone E., Rosano C., Santambrogio F., Curnis F., Corti A., Magni F.,
 RA Siccardi A.G., Padanelli G., Losso R., Aprea B., Bolognesi M.,
 RA Sidoli A., Arosio P.;
 RT "Biochemical characterization and crystal structure of a recombinant
 hen avidin and its acidic mutant expressed in *Escherichia coli*.";
 RL Eur. J. Biochem. 256:453-460(1998).
 CC -!- FUNCTION: THE BIOLOGICAL FUNCTION OF AVIDIN IS NOT KNOWN. FORMS A
 CC STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE MOLECULE OF
 CC BIOTIN PER SUBUNIT OF AVIDIN).
 CC -!- SUBUNIT: HOMOTETRAMER.
 CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN HEN OVIDUCT AND CONCENTRATED IN
 CC EGG WHITE (WHERE IT REPRESENT 0.05% OF THE TOTAL PROTEIN).
 CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
 CC -!- DATABASE: NAME=worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/AV.html".
 CC -!- DATABASE: NAME=prozyme technical fact sheet;
 CC WWW="http://www.prozyme.com/technical/av10data.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X05343; CAA28954.1; -;
 CC EMBL; L27818; AAB59733.1; -;
 CC FIR; A03160; VICH.
 CC FIR; A27518; A27518.
 CC PIR; S11540; S11540.
 CC PDB; 2AVT; 15-JUL-93.
 CC PDB; 1AVD; 31-JAN-94.
 CC PDB; 1AVE; 31-JAN-94.
 CC PDB; 1RAV; 15-JUL-98.
 CC PDB; 2CAM; 15-JUL-98.
 CC InterPro; IPR000088; Avidin.
 CC Pfam; pf01382; Avidin; 1.
 CC PRINTS; PR00709; AVIDIN.
 CC ProDom; PD016055; Avidin; 1.
 CC PROSITE; PS00577; AVIDIN; 1.
 CC Glycoprotein; Signal; Biotin; 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 152 AVIDIN.
 FT DISULFID 28 107 N-LINKED (GLCNAC...).
 FT CARBOHYD 41 41 BIOTIN (NON-COVALENT).
 FT BINDING 57 57 I -> T (IN APPROX. 50% OF THE CHAINS).
 FT VARIANT 58 58 G -> S (IN REF. 3).
 FT CONFLICT 22 22 E -> Q (IN REF. 2 AND 3).
 FT CONFLICT 77 77
 FT STRAND 32 36
 FT TURN 37 38
 FT STRAND 41 44
 FT TURN 49 50

FT STRAND 54 58
 FT STRAND 71 74
 FT STRAND 77 77
 FT TURN 83 84
 FT STRAND 87 93
 FT STRAND 100 109
 FT STRAND 115 124
 FT HELIX 130 135
 FT STRAND 137 146
 SQ SEQUENCE 152 AA; 16769 MW; 1D55A4491D5BFD5C CRC64;

 Query Match 80.6%; Score 681; DB 1; Length 152;
 Best Local Similarity 97.0%; Pred. No. 5.2e-61;
 Matches 128; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 30 PEIPARKGSLCKWTKNDLGSNMTCGAVNSRGFTGTGTYITAVTATSNKEIKESPLHGTQNTI 89
 DB 21 PCLSAKCSLTKGKWNIDLGSNMTICAVNSRGFTGTGTYITAVTATSNKEIKESPLHGTQNTI 80
 QY 90 NKRTOPTFGFTVNWKFSESTTVFTGQCFIDRNGKEVLTMTLLRSVNDIGDDWKATRVG 149
 DB 81 NKRTOPTFGFTVNWKFSESTTVFTGQCFIDRNGKEVLTMTLLRSVNDIGDDWKATRVG 140
 QY 150 INIFRLRTPQKE 161
 DB 141 INIFRLRTPQKE 152

 RESULT 2
 AVRA_CHICK
 ID AVRA_CHICK STANDARD; PRT; 150 AA.
 AC P56734;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Avidin-related protein 4/5 precursor.
 GN AVRA AND AVRS.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEHORN; TISSUE=Oviduct;
 RX MEDLINE=94170814; PubMed=8125122;
 RA Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
 RA Toimela T.A., Helenius M.A., Kulomaa M.S.;
 RT "Molecular cloning and nucleotide sequence of chicken avidin-related
 genes 1-5.";
 RL Eur. J. Biochem. 220:615-621(1994).
 CC -!- MISCELLANEOUS: THE SEQUENCES OF THE CODING REGIONS OF GENES AVRA
 CC AND AVRS ARE IDENTICAL.
 CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z22883; -; NOT_ANNOTATED_CDS.
 CC HSSP; P02701; 1RAV.
 CC InterPro; IPR000088; Avidin.
 CC Pfam; PF01382; Avidin; 1.
 CC PRINTS; PR00709; AVIDIN.
 CC ProDom; PD016055; Avidin; 1.
 CC PROSITE; PS00577; AVIDIN; 1.
 CC Biotin; Signal; Multigene family.
 FT SIGNAL 1 24
 FT CHAIN 1 24 POTENTIAL.

```
FT CHAIN 25 150 AVIDIN-RELATED PROTEIN 4/5.
FT BINDING 57 57 BIOTIN (NON-COVALENT) (BY SIMILARITY).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 150 AA; 16644 MW; 9A6C6C6310EFEL3A CRC64;

Query Match 59.1%; Score 499; DB 1; Length 150;
Best Local Similarity 75.0%; Pred. No. 7.2e-43;
Matches 99; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 30 PEIPARKCSLTGKWTNDLGSNMNTIGAVNSRGEGFTGYITAVTATSNKIKESPLHGTONTI 89
Db 21 PSLARKCSLTGKWTNNLGSIMTIRAVNSRGEGFTGYLTAVADNPGNIKLSPLLGIOH-- 78
QY 90 NKRTQPTFGFTVNNKFSSESTVFTGQCFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
Db 79 KRASQPTFGFTVHWNFSESTVFTGQCFIDRNGKEVLKTMWLLRSSVNDISYDWKATRVG 138
QY 150 INIFTRLTQKE 161
Db 139 YNNFTRLCTVEE 150

RESULT 3
AVR3_CHICK
ID AVR3_CHICK STANDARD; PRT; 150 AA.
AC P56733;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Avidin-related protein 3 precursor.
GN AVR3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN; TISSUE=Oviduct;
RX MEDLINE=94170814; PubMed=8125122;
RA Keinonen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
RA Tolmela T.A., Helenius M.A., Kulomaa M.S.;
RT "Molecular cloning and nucleotide sequence of chicken avidin-related
genes 1-5."
RL Eur. J. Biochem. 220:615-621(1994).
CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AJ237659; CAB39894.1; -
CC HSSP; P02701; 1RAV.
CC InterPro; IPR000088; Avidin.
CC Pfam; PF01382; Avidin; 1.
CC PRINTS; PR00709; AVIDIN.
CC ProDom; PD016055; Avidin; 1.
CC PROSITE; PS00577; AVIDIN; 1.
CC Biotin; Signal; Multigene family.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 150 AVIDIN-RELATED PROTEIN 3.
CC BINDING 57 57 BIOTIN (NON-COVALENT) (BY SIMILARITY).
CC CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 150 AA; 16657 MW; 84DC03926A6BE21C CRC64;
```

```
Query Match 55.5%; Score 469; DB 1; Length 150;
Best Local Similarity 69.7%; Pred. No. 7e-40;
Matches 92; Conservative 11; Mismatches 27; Indels 2; Gaps 1;

QY 30 PEIPARKCSLTGKWTNDLGSNMNTIGAVNSRGEGFTGYITAVTATSNKIKESPLHGTONTI 89
Db 21 PSLARKCSLTGKWTNNLGSIMTIRAVNSRGEGFTGYLTAVADNPGNIKLSPLLGIOH-- 78
QY 90 NKRTQPTFGFTVNNKFSSESTVFTGQCFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
Db 79 KRASQPTFGFTVHWNFSESTVFTGQCFIDRNGKEVLKTMWLLRSSVNDISYDWKATRVG 138
QY 150 INIFTRLTQKE 161
Db 139 YNNFTRLCTVEE 150

RESULT 4
AVR7_CHICK
ID AVR7_CHICK STANDARD; PRT; 150 AA.
AC P56736; Q9W6V4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Avidin-related protein 7 precursor.
GN AVR7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RHODE ISLAND;
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA Kulomaa M.S.;
RT "Characterization and chromosomal localization of the chicken avidin
gene family."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AJ237659; CAB39894.1; -
CC HSSP; P02701; 1RAV.
CC InterPro; IPR000088; Avidin.
CC Pfam; PF01382; Avidin; 1.
CC PRINTS; PR00709; AVIDIN.
CC ProDom; PD016055; Avidin; 1.
CC PROSITE; PS00577; AVIDIN; 1.
CC Biotin; Signal; Multigene family.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 150 AVIDIN-RELATED PROTEIN 7.
CC BINDING 57 57 BIOTIN (NON-COVALENT) (BY SIMILARITY).
CC CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 150 AA; 16543 MW; 30F335611F65923D CRC64;

Query Match 55.3%; Score 467; DB 1; Length 150;
Best Local Similarity 68.9%; Pred. No. 1.1e-39;
Matches 91; Conservative 11; Mismatches 28; Indels 2; Gaps 1;
```



```
Qy 150 INIFTRLRTQKE 161
      | | | | | : |
```

Query Match	18.2%;	Score 154;	DB 1;	Length 107;
Best Local Similarity	63.5%;	Pred. No. 1.2e-08;		

[illegible]

```

RESULT 11
SAVL_STRLV          STANDARD;          PRT;    183 AA.
AC Q53532;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Streptavidin VI precursor (SA VI).
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1936;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95359204; PubMed=7632734;
RA Bayer E.A., Kulik T., Adar R., Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
from Streptomyces."
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -!- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC -!- FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
SR EMBL; S78777; AAB35015.1; -.
DR HSSP; P22629; LSRI.
DR InterPro; IPR000088; Avidin.
DR Pfam; PF01382; Avidin; 1.
DR ProDom; PD016055; Avidin; 1.
DR PROSITE; PS00577; AVIDIN; 1.
KW Signal; Biotin.
FT SIGNAL          1      24      BY SIMILARITY.
FT CHAIN           25     183     INVOLVED IN BIOTIN BINDING (BY
FT BINDING         67     67     SIMILARITY).
FT BINDING         78     78     INVOLVED IN BIOTIN BINDING (BY
FT BINDING         116    116     SIMILARITY).
FT BINDING         132    132     INVOLVED IN BIOTIN BINDING (BY
FT BINDING         144    144     INVOLVED IN BIOTIN BINDING (BY
FT BINDING         183    186    INVOLVED IN BIOTIN BINDING (BY
FT SEQUENCE        183 AA; 18864 MW; 4652D8BB14847DE0 CRC64;

Query Match          16.9%; Score 142.5; DB 1; Length 183;
Best Local Similarity 32.3%; Pred. No. 3.2e-07;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

QY 31 EIPARKSLTGKWNLDGSMWITGAVNSRGFTGTITAV-TATSNKEIKESPLHGTONTI 89
DB 33 QVSAEAGITGTYNOLGSGFTFIVA-GADGALTGTYESAVGNRESYVLTRGYDSPATD 91
QY 90 NKRTOPTFGFTYNNK-----FSESTTVFTGCQFIDRNGKEV-LKTMILLRSSVNDIGDWW 144
DB 92 GSST--ALGWTVAWKNRYNAHSATTSQGV---GGTEARINTQWLLTSGTTE-ANAWK 145
QY 145 AFRVGINIFTRLR 157
DB 146 STLVGHDFTTKVK 158

```

```

RESULT 12
SAV_STRAV          STANDARD;          PRT;    183 AA.
AC P22629;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Streptavidin precursor.
OS Streptomyces avidinii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1895;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-64.
RX MEDLINE=86148514; PubMed=3951999;
RA Argarana C.E., Kuntz I.D., Birken S., Axel R., Cantor C.R.;
RT "Molecular cloning and nucleotide sequence of the streptavidin gene."
RL Nucleic Acids Res. 14:1871-1882(1986).
RN [2]
RP IMPORTANCE OF TRP IN BIOTIN-BINDING.
RX MEDLINE=89134083; PubMed=3223904;
RA Gitlin G., Bayer E.A., Wilchek M.;
RT "Studies on the biotin-binding site of streptavidin. Tryptophan
residues involved in the active site."
RL Biochem. J. 256:279-282(1988).
RN [3]
RP IMPORTANCE OF TYR IN BIOTIN-BINDING.
RX MEDLINE=90351377; PubMed=2386489;
RA Gitlin G., Bayer E.A., Wilchek M.;
RT "Studies on the biotin-binding sites of avidin and streptavidin.
Tyrosine residues are involved in the binding site."
RL Biochem. J. 269:527-530(1990).
RN [4]
RP CELL-BINDING.
RX MEDLINE=90358825; PubMed=2390089;
RA Alon R., Bayer E.A., Wilchek M.;
RT "Streptavidin contains an RYD sequence which mimics the RGD receptor
domain of fibronectin."
RL Biochem. Biophys. Res. Commun. 170:1236-1241(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.
RX MEDLINE=89100243; PubMed=2911722;
RA Weber P.C., Ohlendorf D.H., Wendoloski J.J., Salem F.R.;
RT "Structural origins of high-affinity biotin binding to streptavidin."
RL Science 243:85-88(1989).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.
RX MEDLINE=97337436; PubMed=9194176;
RA Freitag S., le Trong I., Klumb L., Stayton P.S., Stenkamp R.E.;
RT "Structural studies of the streptavidin binding loop."
RL Protein Sci. 6:1157-1166(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS) OF 37-157.
RX MEDLINE=97294734; PubMed=9148939;
RA Katz B.A., Cass R.T.;
RT "In crystals of complexes of streptavidin with peptide ligands
containing the HPQ sequence the pKa of the peptide histidine is less
than 3.0."
RL J. Biol. Chem. 272:13220-13228(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 37-157.
RX MEDLINE=98070615; PubMed=9405158;
RA Katz B.A.;
RT "Binding of biotin to streptavidin stabilizes intersubunit salt
bridges between Asp61 and His87 at low pH."
RL J. Mol. Biol. 274:776-800(1997).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 37-157 OF MUTANTS.
RX MEDLINE=98300340; PubMed=9636711;
RA Freitag S., le Trong I., Chilkoti A., Klumb L.A., Stayton P.S.,
RA Stenkamp R.E.;
RT "Structural studies of binding site tryptophan mutants in the high-
affinity streptavidin-biotin complex."

```



```

FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 696 705 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 750 761 BY SIMILARITY.
FT DISULFID 755 770 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 793 808 BY SIMILARITY.
FT DISULFID 810 819 BY SIMILARITY.
FT DISULFID 826 837 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 848 857 BY SIMILARITY.
FT DISULFID 864 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 902 913 BY SIMILARITY.
FT DISULFID 907 922 BY SIMILARITY.
FT DISULFID 924 933 BY SIMILARITY.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 477 780 MISSING (IN ISOFORM 1B).
FT CONFLICT 279 279 L -> S (IN REF. 2).
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;

Query Match 16.6%; Score 140; DB 1; Length 1064;
Best Local Similarity 33.6%; Pred. NO. 5.2e-06;
Matches 42; Conservative 16; Mismatches 51; Indels 16; Gaps 4;

QY 37 CSLTCKWINDLGSNMTICAVNSRCGFTCTYITAVTATSNELKESPLHGTQNTI-----N 90
DB 939 CDLEGHMINECDQYTI-TKSTGNMLGDYTI-----NTERALGYAAFTVYVGYASN 909
QY 91 KRQTPPTVNVRFSESTVFTQCFDRNGKEVLRKPMWILRSSVNDIGDDWKATRVGI 150
DB 990 NYDFPSGFTVVRDNGOSTTSWTGCHL-CDGEELYITWTWNTNMVSTCODIKSNMVGQ 1048
QY 151 NIFTR 155
DB 1049 DKWTR 1053

RESULT 14
SAV2_STRVL
-ID SAV2_STRVL STANDARD; PRT; 183 AA.

```

```

AC Q53533;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Streptavidin V2 precursor (SA V2)
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID-1936;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95359204; PubMed-7632734;
RA Bayer E.A.; Kulik T.; Adair R.; Wilchek M.;
RT "Close similarity among streptavidin-like, biotin-binding proteins
RL Biochim. Biophys. Acta 1263:60-66(1995).
CC -!- FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
CC MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; S78782; AAB35016.1; -.
DR HSP; P22629; IPTS.
DR InterPro; IPR000088; Avidin.
DR Pfam; pf01382; Avidin; 1.
DR PRODOM; PD016055; Avidin; 1.
DR PROSITE; PS00577; AVIDIN; 1.
KW signal; Biotin.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 183 STREPTAVIDIN V2.
FT BINDING 67 67 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 78 78 SIMILARITY).
FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 183 183 INVOLVED IN BIOTIN BINDING (BY
FT SEQUENCE 183 AA; 18833 MW; FEAFFDFEA4ECCA CRC64;

Query Match 16.5%; Score 139.5; DB 1; Length 183;
Best Local Similarity 33.6%; Pred. NO. 6.4e-07;
Matches 43; Conservative 17; Mismatches 55; Indels 13; Gaps 7;

QY 36 KCSLTCKWINDLGSNMTICAVNSRCGFTCTYITAV-TATSNELKESPLHGTQNTINKRTO 94
DB 30 EAGITGTWYNQJGSTFTIYTA-NADGSLTGTYESAVNGESRYVLTGRYDSAFATDGGST- 95
QY 95 PRFGFTVMRK----FSESTVFTQCFDRNGKEV-LRPMWILRSSVNDIGDDWKATRVGI 149
DB 96 -ALGWTVAWKNNYRNAHSATTSWGGYVA---GSEARINTQWLLTSGTT-AANAWKSTLVG 150
QY 150 NIFTR 157
DB 151 HDTFTKVK 158

RESULT 15
ICIL_LYCPE
ID ICIL_LYCPE STANDARD; PRT; 111 AA.

```

```
AC P16231;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-AUG-1992 (Rel. 23, Last annotation update).
DE Wound-induced proteinase inhibitor I precursor.
OS Lycopersicon peruvianum (Peruvian tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4082;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90036835; PubMed=2808345;
RA Wingate V.P.M., Broadway R.M., Ryan C.A.;
RT "Isolation and characterization of a novel, developmentally regulated
RT proteinase inhibitor I protein and cDNA from the fruit of a wild
RT species of tomato.";
RL J. Biol. Chem. 264:17734-17738(1989).
CC -!- SIMILARITY: BELONGS TO THE POTATO CHYMOTRYPSIN INHIBITOR I FAMILY
CC OF SERINE PROTEASE INHIBITOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05094; AAA34180.1; -
DR PIR; A34359; A34359.
DR HSSP; P19873; LHYM.
DR InterPro; IPR000864; Potato_inhibit.
DR Pfam; PF00280; potato_inhibit; 1.
DR PRINTS; PR00292; POTATOINHBT.
DR ProDom; PD002604; Potato_inhibit; 1.
DR PROSITE; PS00285; POTATO_INHIBITOR; 1.
KW Serine protease inhibitor; Signal.
FT SIGNAL 1 23
FT PROPEP 24 36
FT CHAIN 37 111 WOUND-INDUCED PROTEINASE INHIBITOR I.
FT ACT_SITE 87 88 REACTIVE BOND (BY SIMILARITY).
SQ SEQUENCE 111 AA; 12610 MW; 3777F7C63231C265 CRC64;
```

```
Query Match 16.1%; Score 136; DB 1; Length 111;
Best Local Similarity 81.2%; Pred. No. 7.6e-07;
Matches 26; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 1 MESKFAHIIIVFLLATPFETLLARKESDGPFI 32
   |||||||:||||| ||||:|||||
DB 1 MEAKFAHIIIFLLAFSFETLMARKESDGPV 32
```

Search completed: September 17, 2002, 15:43:39
Job time: 125 sec

A:Reference number: A51622; PDB:1AVD
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms, with biotin, residues 27-3
 R:Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M.
 submitted to the Brookhaven Protein Data Bank, March 1993
 A:Reference number: A51623; PDB:1AVE
 R:Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M.
 J. Mol. Biol. 231, 699-710, 1993
 A:Title: Three-dimensional structure of the tetragonal crystal form of egg-white avidin
 A:Reference number: A54974; MUID:93294833
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms
 C:Genetics:
 A:Introns: 27/3; 96/1; 136/2
 C:Superfamily: avidin
 C:Keywords: glycoprotein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-152/Product: avidin #status experimental <MAT>
 F:28-107/Disulfide bonds: #status experimental
 F:41/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 80.9%; Score 684; DB 1; Length 152;
 Best Local Similarity 97.7%; Pred. No. 7.3e-60;
 Matches 129; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 30 PEIPARKCSLTGKWTNDLGSNMITGAVNSRGEFTCTYITAVTATATNEIKESPLHGQTNTI 89
 I : |||||
 DB 21 PLSARKCSLTGKWTNDLGSNMITGAVNSRGEFTCTYITAVTATATNEIKESPLHGQTNTI 80
 I : |||||

QY 90 NKRTQPTFGFTVNNKFSSTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
 I : |||||
 DB 81 NKRPQTFGFTVNNKFSSTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 140
 I : |||||

QY 150 INIFTRLRTOKE 161
 I : |||||
 DB 141 INIFTRLRTOKE 152
 I : |||||

RESULT 2
 S42204
 A:Title: avidin-related protein 4/5 precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
 C:Accession: S42204; S42205
 R:Keinaenen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Helenius,
 Eur. J. Biochem. 220, 615-621, 1994
 A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-5.
 A:Reference number: S42201; MUID:94170614
 A:Accession: S42204
 A:Molecule type: DNA
 A:Residues: 1-150 <KEI>
 A:Cross-references: EMBL:Z22883; NID:q311811
 A:Experimental source: strain White Leghorn; tissue oviduct
 A:Genetics: CH1
 A:Accession: S42205
 A:Molecule type: DNA
 A:Residues: 1-150 <KEW>
 A:Cross-references: EMBL:Z22882; NID:q311812
 A:Experimental source: strain White Leghorn; tissue oviduct
 A:Genetics: CH2
 A:Gene: avr4
 A:Introns: 27/3; 96/1; 136/2
 C:Genetics: <CH2>
 A:Gene: avr5
 A:Introns: 27/3; 96/1; 136/2
 C:Superfamily: avidin
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-150/Product: avidin-related protein 4/5 #status predicted <MAT>
 F:28-105/Disulfide bonds: #status predicted

Query Match 59.1%; Score 499; DB 2; Length 150;

Best Local Similarity 75.0%; Pred. NO. 1e-41;
 Matches 99; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 30 PEIPARKCSLTGKWTNDLGSNMITGAVNSRGEFTCTYITAVTATATNEIKESPLHGQTNTI 89
 I : |||||
 DB 21 PLSARKCSLTGKWTNDLGSNMITGAVNSRGEFTCTYITAVTADNPGNTILSLGLGH-- 78
 I : |||||

QY 90 NKRTQPTFGFTVNNKFSSTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
 I : |||||
 DB 79 KRASQPTFGFTVNNKFSSTVFTGQCIFDRNGKEVLKTMWLLRSSVNDISYDWKATRVG 138
 I : |||||

QY 150 INIFTRLRTOKE 161
 I : |||||
 DB 139 YNNFTRLCTVEE 150
 I : |||||

RESULT 3
 S42203
 A:Title: avidin-related protein 3 precursor - chicken
 N:Alternate names: avr3 protelin
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
 C:Accession: S42203; S39800
 R:Keinaenen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Heleni
 Eur. J. Biochem. 220, 615-621, 1994
 A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-
 A:Reference number: S42201; MUID:94170814
 A:Accession: S42203
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-150 <KEI>
 A:Cross-references: EMBL:Z21612; NID:g65432
 R:Kunnaas, T.A.; Wallen, M.J.; Kulomaa, M.S.
 Biochim. Biophys. Acta 1216, 441-445, 1993
 A:Title: Induction of chicken avidin and related mRNAs after bacterial infection.
 A:Reference number: S39799; MUID:94092737
 A:Accession: S39800
 A:Status: preliminary; translation not shown
 A:Molecule type: mRNA
 A:Residues: 71-150 <KUN>
 A:Cross-references: EMBL:Z21536; NID:g65429
 C:Genetics:
 A:Gene: avr3
 A:Introns: 27/3; 96/1; 136/2
 C:Superfamily: avidin
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-150/Product: avidin-related protein 3 #status predicted <MAT>
 F:28-105/Disulfide bonds: #status predicted

Query Match 55.5%; Score 469; DB 2; Length 150;
 Best Local Similarity 69.7%; Pred. No. 9.1e-39;
 Matches 92; Conservative 11; Mismatches 27; Indels 2; Gaps 1;

QY 30 PEIPARKCSLTGKWTNDLGSNMITGAVNSRGEFTCTYITAVTATATNEIKESPLHGQTNTI 89
 I : |||||
 DB 21 PLSARKCSLTGKWTNDLGSNMITGAVNSRGEFTCTYITAVTADNPGNIKLSPLGLGH-- 78
 I : |||||

QY 90 NKRTQPTFGFTVNNKFSSTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
 I : |||||
 DB 79 KRACOPTFGFTVNNKFSSTVFTGQCIFDRSCKEVLKTKWLORLAVDDISDDWKATRVG 138
 I : |||||

QY 150 INIFTRLRTOKE 161
 I : |||||
 DB 139 YNNFTRLCTVEE 150
 I : |||||

RESULT 4
 S42201
 A:Title: avidin-related protein 1 precursor - chicken
 N:Alternate names: avr1
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999

	Matches	87;	Conservative	12;	Mismatches	31;	Indels	2;	Gaps
QY	30	PEIPARCSLTGKWTNDLGNMTIGAVNSGEFTGYITAVTATSNIEIKESPLHGQTNTI	89						
	I :							:	
Db	21	PSLSARCSTLGWDNDLGSIMTIGAVNDNGEDGYITAVADPNGNITLSPLLGIQH--	78						
	I :							:	
QY	90	NKRTPPTFGTVNKKFSESTVTFGGCFIDRNGKEVLKTNMMLRSSVDIGDDMKATRVG	149						
	:							:	
Db	79	KRASQPTFGTVHNWFSESTSVFVGQCFCVDRSGKEVLKTWLQRQLAYDDISDDWIAIRVG	138						
	I :							:	
QY	150	INIFTRLRTOKE	161						
	I	:							
Db	139	NNDETROHVEE	150						
	I	:							
RESULT	6								
S06251		proteinase inhibitor I precursor - potato (cv. Russet Burbank)							
C:	Species:	Solanum tuberosum (potato)							
A:	Variety:	cv. Russet Burbank							
C:	Date:	28-Feb-1990 #sequence_revision 19-Apr-1996 #text_change 16-Jul-1999							
C:	Accession:	S06251; S060563							
R:	Cleveland, T.E.; Thornburg, R.W.; Ryan, C.A.								
	Plant Mol. Biol.	8, 199-207, 1987							
A:	Title:	Molecular characterization of a wound-inducible inhibitor I gene							
A:	Reference number:	S06251							
A:	Accession:	S06251							
A:	Molecule type:	DNA							
A:	Residues:	1-107 <CLE>							
A:	Cross-references:	EMBL:M17108; NID:g169492; PIDN:AAA33816.1; PID:g169493							
A:	Experimental source:	cultivar Russet Burbank							
A:	Note:	the authors translated the codon CAA for residue 45 as Glu, TCT fo							
R:	Seob, L.J.; Yang, Y.S.								
	Korean J. Bot.	33, 303-307, 1990							
A:	Title:	Nucleotide sequence of a truncated proteinase inhibitor I gene of							
A:	Reference number:	S060563							
A:	Accession:	S060563							
A:	Molecule type:	DNA							
A:	Residues:	1-14 <SEO>							
A:	Cross-references:	EMBL:Z12819							
A:	Experimental source:	cultivar Russet Burbank							
C:	Superfamily:	eglin C							
C:	Keywords:	serine proteinase inhibitor							
F:	1-23/Domain:	signal sequence #status predicted <SIG>							
F:	24-36/Domain:	propeptide #status predicted <PRO>							
F:	37-107/Product:	proteinase inhibitor I #status predicted <MAT>							
	Query Match	18.2%	Score 154;	DB 2;	Length 107;				
	Best Local Similarity	63.5%;	Pred. No. 4.9e-08;						
	Matches	33;	Conservative	5;	Mismatches	10;	Indels	4;	Gaps
QY	1	MESFAHIIVFFLLATPFELLARKESDGPFI----	PARKCSLTGWKWNLDLG	48					
	I				:	:	:	:	
Db	1	MESFAHIIVFFLLATSFELLARKESDGPFEVILOKEFCFCNGKQRPWELIG	52						
	I				:	:	:	:	
RESULT	7								
S25159		proteinase inhibitor I - potato (cultivar Ilam Hardy)							
C:	Species:	Solanum tuberosum (potato)							
A:	Variety:	cultivar Ilam Hardy							
C:	Date:	20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 16-Jul-1999							
C:	Accession:	S25159; S25470							
R:	Beuning, L.B.; Christeller, J.T.								
	submitted to the EMBL Data Library, August 1992								
A:	Description:	Nucleotide seq of a cDNA encoding the proteinase inhibitor							
A:	Reference number:	S25159							
A:	Accession:	S25159							
A:	Molecule type:	mRNA							
A:	Residues:	1-107 <BEU>							
A:	Cross-references:	EMBL:X67675; NID:g21527; PIDN:CAA47907.1; PID:g21528;							


```
F;180-211/Domain: EGF homology <EG02>  
F;218-249/Domain: EGF homology <EG03>  
F;256-287/Domain: EGF homology <EG04>  
F;294-325/Domain: EGF homology <EG05>  
F;332-363/Domain: EGF homology <EG06>  
F;370-401/Domain: EGF homology <EG07>  
F;408-439/Domain: EGF homology <EG08>  
F;446-477/Domain: EGF homology <EG09>  
F;484-515/Domain: EGF homology <EG10>  
F;522-553/Domain: EGF homology <EG11>  
F;560-591/Domain: EGF homology <EG12>  
F;598-629/Domain: EGF homology <EG13>  
F;636-667/Domain: EGF homology <EG14>  
F;674-705/Domain: EGF homology <EG15>  
F;712-743/Domain: EGF homology <EG16>  
F;750-781/Domain: EGF homology <EG17>  
F;788-819/Domain: EGF homology <EG18>  
F;826-857/Domain: EGF homology <EG19>  
F;864-895/Domain: EGF homology <EG20>  
F;902-933/Domain: EGF homology <EG21>  
F;936-1064/Region: avidin-like  
F;23-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,261  
57_451-466,468-477,484-495/Disulfide bonds: #status predicted  
F;489-504,508-515,522-533,527-542,544-553,560-571,565-580,582-591,598-609,603-618,620  
08,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/DI
```

```
Query Match          16.6%; Score 140; DB 2; Length 1064;  
Best Local Similarity 33.6%; Pred. No. 2.1e-05;  
Matches 42; Conservative 16; Mismatches 51; Indels 16; Gaps 4;
```

```
QY      37 CSLSGTGWNTDLGSNMITIGAVNSRGFTGYTATVTSNEIKESPLHGQTNTI-----N 90  
         | | | | : || | | | | | | | | | | | | | | |  
Db       939 CDLEGWMYNECNDQVII-TKTSTGMMLGDYMT-----YNERALGYAAPT VVGVSAN 989
```

```
QY      91 KRTPQTFGFTVNWKFSSESTVTFTGCIFIDRNGKEVLKTMLLRSSVNDIGDDKKATRVGI 150  
         |:||||| :|||:|||| : |:|||| : |: | | | | : |:  
Db       990 NYDFPSFGFTVVDRNQGSTTSWTQCCHL-CDGEVLYTTINTVMYSTCODIKKSNNVGQ 1048
```

```
QY      151 NIFTR 155  
         | | | | : || |
```

```
Db       1049 DKWTR 1053
```

```
RESULT 13  
S57285  
streptavidin v2 precursor - Streptomyces venezuelae  
C:Species: Streptomyces venezuelae  
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999  
C:Accession: S57285  
R:Bayer, E.A.; Kulik, T.; Adar, R.; Wilchek, M.  
Biochim. Biophys. Acta 1263, 60-66, 1995  
A>Title: Close similarity among streptavidin-like, biotin-binding proteins from Strep  
A:Reference number: S57284; MUID:95J59204  
A:Accession: S57285  
A:Molecule type: DNA  
A:Residues: 1-183 <BAY>  
A:Experimental source: strain Tue2605  
C:Superfamily: streptavidin  
F;1-24/Domain: signal sequence #status predicted <SIG>  
F;25-183/Product: streptavidin v2 #status predicted <MAT>
```

```
Query Match          16.5%; Score 139.5; DB 2; Length 183;  
Best Local Similarity 33.6%; Pred. No. 2.6e-06;  
Matches 43; Conservative 17; Mismatches 55; Indels 13; Gaps 7;
```

```
QY      36 KCSLTGKWNTDLGSNMITIGAVNSRGFTGYTIYTA-VATSNEIKESPLHGTQNTINKRTQ 94  
         : || | | || | | : | | | | | | | | | | | | | | |  
Db       38 EAGITGTYWNQLSGTFIVTA-NADGSLTGTHYESAVGNAESRYVL TGTDYSAPATDGGSGT- 95
```

```
QY      95 PTFGTFNVNK-----FSESTTVFTTCQCFIDRNGKEV-LKTMLLRSSVNDIGDDWKATRVC 149  
         |:||| || | | | | | | | | | | | | | | | | | | | |
```

```
db 96 -ALGQVAVKNNVNAHSATWSSGOVVA--GSEARINTQWLLTSGIT-AANAWKSTLYG 150
Qy 150 INIFTRLR 157
      : ||:::
db 151 HDTFTVKY 158
```

```

RESULT 14
A34359
trypsin inhibitor I precursor - peruvian tomato
C:Species: Lycopersicon peruvianum (Peruvian tomato)
C:Date: 15-JUN-1990 #sequence_revision 15-Jun-1990 #text_change 16-Jul-1999
C:Accession: A39547; A34359

```

R; Wingate, V.P.M.; Ryan, C.A.
J. Biol. Chem. 266, 5814-5818, 1991
A; Title: A novel fruit-expressed trypsin inhibitor I gene from a wild species of tomato
A; Reference number: A39547; MUID: 91170261

A:Accession: A39547
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-111 <WIN>
A:Cross-references: GR:M59427
A:Note: the authors translated the codon AAA for residue 37 as Ile, GAC for residue 42 as Lys, and GAG for residue 43 as Glu.
R:Winkler, V.P.M.; Broadway, R.M.; Ryan, C.A.

J. Biol. Chem. 264, 17734-17738, 1989

A: Accession: A34359
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-111 (W12)
A: Cross-references: GB: J05094; NID: g170476; PIDN: AAA34180.1; PID: g170477
C: Superfamily: eglin C

Query Match 16.1%; Score 136; DB 2; Length 111;
 Best Local Similarity 81.2%; Pred. No. 3e-06;
 Matches 26; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

RESULT 15
A48836
fibropellin C precursor - sea urchin (Strongylocentrotus purpuratus)
N;Alternate names: ebf repeat-containing protein; epidermal growth factor-related protein
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 01-Dec-1993 #sequence_revision 10-Nov-1994 #text_change 16-Jul-1999
C;Accession: A48836

Dev. Biol. 157, 526-538, 1993

A:Accession: A48836
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-570 <BIS>
A:Cross-references: GB:L07045; NID:g3l0659; PID:g3l0660
A:Note: sequence extracted from NCBI backbone (NCBIN:132724 NCBIID:132725)

C:Superfamily: C1r/C1s repeat homology; EGF homology
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-370/Product: fibronectin c #status predicted <FIB>
F:57-175/Domain: EGF homology <EG1>
F:176-211/Domain: EGF homology <EG2>
F:214-249/Domain: EGF homology <EG3>
F:252-287/Domain: EGF homology <EG4>
F:290-325/Domain: EGF homology <EG5>
F:328-363/Domain: EGF homology <EG6>
F:366-401/Domain: EGF homology <EG7>
F:404-439/Domain: EGF homology <EG8>

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 15:39:39 ; Search time 12.95 Seconds
(without alignments)
303.670 Million cell updates/sec

Title: US-09-743-690-7

Perfect score: 845

Sequence: 1 MESKFAHIIFFLLATPFT.....DWKATRGVGINIFRLRTQKE 161

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/5A-COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/5B-COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	80.9	152	4 US-08-831-399-4	Sequence 4, Appli
2	142.5	16.9	159	3 US-08-628-540-1	Sequence 1, Appli
3	142.5	16.9	159	3 US-08-948-097-17	Sequence 17, Appl
4	142.5	16.9	159	4 US-08-941-100-1	Sequence 1, Appli
5	142.5	16.9	159	4 US-08-941-100-5	Sequence 5, Appli
6	142.5	16.9	183	4 US-08-831-399-2	Sequence 2, Appli
7	142.5	16.9	183	5 PCT-US93-05240-14	Sequence 14, Appl
8	142.5	16.9	183	6 US-08-831-399-5	Patent No. 5168049
9	142.5	16.9	415	2 US-08-491-988-7	Sequence 7, Appli
10	142.5	16.9	435	2 US-08-491-988-5	Sequence 5, Appli
11	141	16.7	402	2 US-08-491-988-9	Sequence 9, Appli
12	135.5	16.0	128	1 US-08-211-833-2	Sequence 2, Appli
13	135.5	16.0	128	1 US-08-434-718-2	Sequence 2, Appli
14	135.5	16.0	128	4 US-08-831-399-16	Sequence 16, Appl
15	115	13.6	21	1 US-08-554-586-2	Sequence 2, Appli
16	115	13.6	21	1 US-08-554-586-3	Sequence 3, Appli
17	67	7.9	462	2 US-08-752-307B-7	Sequence 7, Appli
18	67	7.9	465	2 US-08-752-307B-5	Sequence 5, Appli
19	66.5	7.9	997	1 US-08-232-540-1	Sequence 1, Appli
20	66.5	7.9	997	1 US-08-428-949A-1	Sequence 1, Appli
21	66.5	7.9	997	1 US-08-428-948A-1	Sequence 1, Appli
22	66.5	7.9	997	2 US-08-428-946-1	Sequence 1, Appli
23	66.5	7.9	997	5 PCT-US95-04656-1	Sequence 1, Appli
24	66.5	7.9	998	1 US-08-233-008A-6	Sequence 6, Appli
25	66.5	7.9	1021	1 US-08-233-008A-2	Sequence 2, Appli
26	66	7.8	682	4 US-08-982-785A-9	Sequence 9, Appli
27	66	7.8	703	4 US-08-910-925-4	Sequence 4, Appli

28 65.5 7.8 1257 1 US-08-340-428B-49 Sequence 49, Appli
29 64.5 7.6 325 1 US-08-292-549-2 Sequence 2, Appli
30 64.5 7.6 325 4 US-09-042-785A-9 Sequence 9, Appli
31 64.5 7.6 325 5 PCT-US91-02207-2 Sequence 2, Appli
32 64.5 7.6 674 1 US-08-317-522A-3 Sequence 3, Appli
33 64.5 7.6 674 1 US-08-439-818A-3 Sequence 3, Appli
34 64.5 7.6 674 2 US-08-751-965-3 Sequence 3, Appli
35 64.5 7.6 674 2 US-08-738-975-3 Sequence 3, Appli
36 64.5 7.6 674 2 US-08-728-626-3 Sequence 3, Appli
37 64.5 7.6 674 3 US-08-808-599A-3 Sequence 3, Appli
38 64.5 7.6 749 1 US-08-317-522A-2 Sequence 2, Appli
39 64.5 7.6 749 1 US-08-439-818A-2 Sequence 2, Appli
40 64.5 7.6 749 2 US-08-751-965-2 Sequence 2, Appli
41 64.5 7.6 749 2 US-08-738-975-2 Sequence 2, Appli
42 64.5 7.6 749 2 US-08-728-626-2 Sequence 2, Appli
43 64.5 7.6 749 3 US-08-808-599A-2 Sequence 2, Appli
44 64.5 7.6 2476 2 US-08-276-967-2 Sequence 2, Appli
45 64 7.6 630 1 US-08-487-890A-113 Sequence 113, App

ALIGNMENTS

RESULT 1
US-08-831-399-4
; Sequence 4, Application US/08831399
; Patent No. 6312916
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Erhard; Muller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Hans; Brandstetter, Hans
; TITLE OF INVENTION: Recombinant Inactive Core
; TITLE OF INVENTION: Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,399
; FILING DATE: 1-April-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 37 718.8
; FILING DATE: 16-September-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6312916man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 1105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 638-3884
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-831-399-4

Query Match 80.9%; Score 684; DB 4; Length 152;
Best Local Similarity 97.7%; Pred. No. 2.2e-73;
Matches 129; Conservative 1; Mismatches 2; Indels 0; Gaps 0;


```
;
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1)
; OTHER INFORMATION: The residue in this position can be any amino
; OTHER INFORMATION: acid.
US-08-941-100-1

Query Match          16.9%; Score 142.5; DB 4; Length 159;
Best Local Similarity 32.3%; Pred. No. 3.3e-09;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

Qy 31 EIPARKCSLTGKWTNDLGSNMITGAVNSRGFTGTITAV-TATSNKESPLHGTONTI 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 QVSAEAGITGTWYNQLGSTFFIVTA-GADGALTGTYESAVGNAESRYVLTGRYDSAPATD 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 90 NKRTOPTFGFTVNWK----FSESTTVFTGQCFFDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 GSGT--ALGWTVAWKNNRYNAHSATWSGQYV---GGAEARINTQWLLTSGTTE-ANAWK 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 145 ATRVGINIFTRLR 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 STLVGHDFTFKVK 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-08-941-100-5
; Sequence 5, Application US/08941100B
; Patent No. 6207390
; GENERAL INFORMATION:
; APPLICANT: Cantor, Charles R.
; TITLE OF INVENTION: Reduced Affinity Streptavidin
; FILE REFERENCE: BU-03165
; CURRENT APPLICATION NUMBER: US/08/941,100B
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 08/469,353
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/420,010
; PRIOR FILING DATE: 1995-04-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Streptomyces avidinii
US-08-941-100-5

Query Match          16.9%; Score 142.5; DB 4; Length 159;
Best Local Similarity 32.3%; Pred. No. 3.3e-09;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

Qy 31 EIPARKCSLTGKWTNDLGSNMITGAVNSRGFTGTITAV-TATSNKESPLHGTONTI 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 9 QVSAEAGITGTWYNQLGSTFFIVTA-GADGALTGTYESAVGNAESRYVLTGRYDSAPATD 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 90 NKRTOPTFGFTVNWK----FSESTTVFTGQCFFDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 GSGT--ALGWTVAWKNNRYNAHSATWSGQYV---GGAEARINTQWLLTSGTTE-ANAWK 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 145 ATRVGINIFTRLR 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 STLVGHDFTFKVK 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-831-399-2
; Sequence 2, Application US/08831399
; Patent No. 6312916
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Erhard; Muller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandstetter, Hans
```

```
;
; TITLE OF INVENTION: Recombinant Inactive Core
; TITLE OF INVENTION: Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,399
; FILING DATE: 1-April-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 37 718.8
; FILING DATE: 16-September-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6312916man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 1105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-831-399-2

Query Match          16.9%; Score 142.5; DB 4; Length 183;
Best Local Similarity 32.3%; Pred. No. 4e-09;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

Qy 31 EIPARKCSLTGKWTNDLGSNMITGAVNSRGFTGTITAV-TATSNKESPLHGTONTI 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 33 QVSAEAGITGTWYNQLGSTFFIVTA-GADGALTGTYESAVGNAESRYVLTGRYDSAPATD 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 90 NKRTOPTFGFTVNWK----FSESTTVFTGQCFFDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 GSGT--ALGWTVAWKNNRYNAHSATWSGQYV---GGAEARINTQWLLTSGTTE-ANAWK 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 145 ATRVGINIFTRLR 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 146 STLVGHDFTFKVK 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
PCT-US93-05240-14
; Sequence 14, Application PC/TUS9305240
; GENERAL INFORMATION:
; APPLICANT: NAGARAJAN, VASANTHA
; TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN FROM BACILLUS
; TITLE OF INVENTION: SUBTILLIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DU PONT COMPANY
; STREET: BARLEY MILL PLAZA 36
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19880-0036
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05240
; FILING DATE: 19930527
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GEIGER, KATHLEEN W
; REFERENCE/DOCKET NUMBER: CR 9029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-2118
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US93-05240-14

Query Match 16.9%; Score 142.5; DB 5; Length 183;
Best Local Similarity 32.3%; Pred. NO. 4e-09;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

QY 31 EIPARKSLTGKWTNDLGSNMWTIGAVNSRGEFTGTITAV-TATSNFKEISPLHGTONTI 89
Db 33 QVSAAEAGITGTWYNLQGSTFIYTA-GADGALTGTYESAVGNAESRYVLTRDYDSAPATD 91
QY 90 NKRTQPTFGFTVNMK----PSESTVTFGTGCFIDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
Db 92 GSGT--ALGHTYAKKNYRNAHSATTWSGOYV---GGAERINTQWLLTSGTTE-ANAMK 145
QY 145 ATRVGINIFRLR 157
Db 146 STLVGHDFTKVK 158

RESULT 8
5168049-5
; Patent No. 5168049
; APPLICANT: MEADE, HARRY M.; GARWIN, JEFFREY L.
; TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN-LIKE
; POLYPEPTIDES
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/185,329
; FILING DATE: 21-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 656,873
; FILING DATE: 02-OCT-1984
; SEQ ID NO: 5
; LENGTH: 183
5168049-5

Query Match 16.9%; Score 142.5; DB 6; Length 183;
Best Local Similarity 32.3%; Pred. NO. 4e-09;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

QY 31 EIPARKSLTGKWTNDLGSNMWTIGAVNSRGEFTGTITAV-TATSNFKEISPLHGTONTI 89
Db 33 QVSAAEAGITGTWYNLQGSTFIYTA-GADGALTGTYESAVGNAESRYVLTRDYDSAPATD 91
QY 90 NKRTQPTFGFTVNMK----PSESTVTFGTGCFIDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
Db 92 GSGT--ALGHTYAKKNYRNAHSATTWSGOYV---GGAERINTQWLLTSGTTE-ANAMK 145
QY 145 ATRVGINIFRLR 157

```

Query Match	16.7%;	Score 141;	DB 2;	Length 402;
Best Local Similarity	33.3%;	Pred. No. 1.9e-08;		
Matches	46;	Conservative 18;	Mismatches 58;	Indels 16;
Gaps				

Qy	29	GPEIPA---RKCSLTGKWTNDLGSNM	TIGAVNSRGEGTGYITAV-TAT	SNEIKESPLHG	84
			:	:	:
Db	267	GLEAPAAAPAEAGITGWYNQLGST	FTVTA-GADGALTGTYESAVGNA	ESRYVLTGRYDS	325
			:	:	:
Qy	85	TQNTINKRTOPTGCTVNMK----	FSESTVFTGCQFIDNRGKGV-L	KTMWLLRSSVNDI	139
			:	:	:
Db	326	APATDGSGT--ALGWTYVAMKNRY	NAHSQGVY---GGAEARINTQW	LLTSGTTE-	379
			:	:	:
Ov	140	GDDWKATRVGINFTFRLR			157

DD 360 ANAWKSILVGHDFINAK 397

RESULT 12
US-08-211-833-2
; Sequence 2, Application US/08211833
; Patent No. 5489528

```

: 03 00 211 003 2
:
: Sequence 2, Application US/09211833
: Patent No. 5489528
:
: GENERAL INFORMATION:
:
: APPLICANT: KOPETZKI, Erhard
: APPLICANT: RUDOLPH, Rainer
: APPLICANT: GROSSMANN, Adelbert
:
: TITLE OF INVENTION: RECOMBINANT CORE STREPTAVIDIN
:
: NUMBER OF SEQUENCES: 6
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
: STREET: 655 Fifteenth Street N.W., Suite 330
:
: CITY: Washington
:
: STATE: D.C.
:
: COUNTRY: USA
:

```

STATE: D.C.
COUNTRY: USA
ZIP: 2005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/211,833
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 41 35 543.1
FILING DATE: 28-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kitts, Monica C.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P564-4013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-211-833-2

Query Match 16.0%; Score 135.5; DB 1; Length 128;
Best Local Similarity 32.8%; Pred. No. 1.6e-08;
Matches 42; Conservative 18; Mismatches 55; Indels 13; Gaps 7;
QY 36 KCSLTCKWTNDLGSNMTCGAVNSRCEFTCTVTAV-TATSNEIKESPLHCTONTINKRTO 94
DB 3 EAGITGTWYNQLGSTFIVTA-GADGALTGTYESAVGNAESRYVLTKGYDSAPATDGSCT- 60
QY 95 PTEGFTVNWK---FSESTTVFTGQCFIDRNGKEV-LKTMWLLRSSVNDIGDDWKATRYG 149
DB 61 -ALGWTVAWKNNYRNAHSATTWSGOYV---GGAERINTQWLLTSGTTE-ANAWKSTLWG 115
QY 150 INIFTRLR 157
DB 116 HDTFTKVK 123

RESULT 13

US-08-434-718-2
Sequence 2, Application US/08434718
Patent No. 5672691
GENERAL INFORMATION:
APPLICANT: KOPETZKI, Erhard
APPLICANT: RUDOLPH, Rainer
APPLICANT: GROSSMANN, Adelbert
TITLE OF INVENTION: RECOMBINANT CORE STREPTAVIDIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nikolaus, Marmelstein, Murray & Oram
STREET: 655 Fifteenth Street N.W., Suite 330
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,718
FILING DATE: 4-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/211,833
FILING DATE: 28-APRIL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 41 35 543.1
FILING DATE: 28-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: KILLS, Monica C.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: F564-5007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-434-718-2

Query Match 16.0%; Score 135.5; DB 1; Length 128;
Best Local Similarity 32.8%; Pred. No. 1.6e-08;
Matches 42; Conservative 18; Mismatches 55; Indels 13; Gaps 7;

QY 36 KCSLTCKWTNDLGSNMTCGAVNSRCEFTCTVTAV-TATSNEIKESPLHCTONTINKRTO 94

DB 3 EAGITGTWYNQLGSTFIVTA-GADGALTGTYESAVGNAESRYVLTKGYDSAPATDGSCT- 60
QY 95 PTEGFTVNWK---FSESTTVFTGQCFIDRNGKEV-LKTMWLLRSSVNDIGDDWKATRYG 149
DB 61 -ALGWTVAWKNNYRNAHSATTWSGOYV---GGAERINTQWLLTSGTTE-ANAWKSTLWG 115
QY 150 INIFTRLR 157
DB 116 HDTFTKVK 123

RESULT 14

US-08-831-399-16
Sequence 16, Application US/08831399
Patent No. 6312916
GENERAL INFORMATION:
APPLICANT: KOPETZKI, Erhard; Muller, Rainer;
APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandstetter, Hans
TITLE OF INVENTION: Recombinant Inactive Core
TITLE OF INVENTION: Streptavidin Mutants
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,399
FILING DATE: 1-APRIL-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 196 13 053.0
FILING DATE: 1-APRIL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 196 37 718.8
FILING DATE: 16-SEPTEMBER-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6312916man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: HUBR 1105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 128 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-831-399-16

Query Match 16.0%; Score 135.5; DB 4; Length 128;
Best Local Similarity 32.8%; Pred. No. 1.6e-08;
Matches 42; Conservative 18; Mismatches 55; Indels 13; Gaps 7;

QY 36 KCSLTCKWTNDLGSNMTCGAVNSRCEFTCTVTAV-TATSNEIKESPLHCTONTINKRTO 94
DB 3 EAGITGTWYNQLGSTFIVTA-GADGALTGTYESAVGNAESRYVLTKGYDSAPATDGSCT- 60
QY 95 PTEGFTVNWK---FSESTTVFTGQCFIDRNGKEV-LKTMWLLRSSVNDIGDDWKATRYG 149
DB 61 -ALGWTVAWKNNYRNAHSATTWSGOYV---GGAERINTQWLLTSGTTE-ANAWKSTLWG 115
QY 150 INIFTRLR 157
DB 116 HDTFTKVK 123

Db 116 HDTFTKVK 123

```
RESULT 15
US-08-554-586-2
; Sequence 2, Application US/08554586
; Patent No. 5767379
; GENERAL INFORMATION:
; APPLICANT: BASZCZYNSKI, Chris
; APPLICANT: HOOD, Elizabeth
; APPLICANT: MADDOCK, Sheila
; APPLICANT: MEYER, Terry Euclaire
; APPLICANT: REGISTER III, James C.
; APPLICANT: WITCHER, Derrick
; APPLICANT: HOWARD, John A.
; TITLE OF INVENTION: COMMERCIAL PRODUCTION OF AVIDIN IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/554,586
; FILING DATE: 06-NOV-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/352/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-554-586-2
```

```
Query Match 13.6%; Score 115; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ARKCSLTGKWTNDLGSNMTIG 54
DB 1 ARKCSLTGKWTNDLGSNMTIG 21
```

Search completed: September 17, 2002, 15:42:24
Job time: 165 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 15:37:39 ; Search time 30.07 Seconds
(without alignments)
594.708 Million cell updates/sec

Title: US-09-743-690-7
Perfect score: 845
Sequence: 1 MESKFAHIIFFLLATPFET.....DWKATRVGINIFRLTRTOKE 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	845	100.0	161	21	AA44699
2	684	80.9	152	18	AAW29307
3	684	80.9	152	20	AA117867
4	681	80.6	152	20	AA30130
5	678	80.2	400	20	AA03098
6	670	79.3	128	22	AA35269
7	656	77.6	128	22	AA35271
8	148	17.5	402	15	AAW56485
9	145.5	17.2	183	19	AAW59217
10	142.5	16.9	159	22	AA35270
11	142.5	16.9	160	10	AAW93531

12	142.5	16.9	161	21	AAW80515	Streptomyces avidi
13	142.5	16.9	162	21	AAW80514	Streptomyces avidi
14	142.5	16.9	163	9	AAW80160	Biosynthetic prote
15	142.5	16.9	182	8	AAW70492	Streptavidin seque
16	142.5	16.9	183	7	AAW60625	Sequence of a stre
17	142.5	16.9	183	10	AAW93530	Streptavidin prote
18	142.5	16.9	183	14	AAW44491	Streptavidin gene.
19	142.5	16.9	183	18	AAW29306	Wild-type streptav
20	142.5	16.9	183	19	AAW59216	S. avidinii strept
21	142.5	16.9	183	20	AAW17868	Streptococcus stre
22	142.5	16.9	183	21	AAW44701	Streptavidin prote
23	142.5	16.9	183	22	AAW30692	Amino acid sequenc
24	142.5	16.9	184	21	AAW80513	Streptomyces avidi
25	142.5	16.9	186	21	AAW80512	Streptomyces avidi
26	142.5	16.9	412	22	AAW30694	A fusion of anti-C
27	142.5	16.9	415	15	AAW56484	ScFV PRAS109 and p
28	142.5	16.9	423	22	AAW30695	A fusion of anti-C
29	142.5	16.9	431	22	AAW30693	A fusion of single
30	142.5	16.9	435	15	AAW56483	ScFV PRAS108 and p
31	140.5	16.6	159	18	AAW29311	Streptavidin prote
32	139.5	16.5	188	21	AAW44700	Potato proteinase
33	139.5	16.5	183	19	AAW59218	S. avidinii strept
34	137.5	16.3	159	18	AAW29310	Streptavidin prote
35	137	16.2	186	21	AAW84021	Amino acid sequenc
36	136.5	16.2	159	18	AAW29309	Streptavidin prote
37	135.5	16.0	128	14	AAW34722	Core streptavidin.
38	135.5	16.0	128	17	AAW04211	Streptomyces avidi
39	135.5	16.0	128	18	AAW29308	Recombinant Core-s
40	135.5	16.0	159	18	AAW29314	Streptavidin prote
41	135.5	16.0	270	20	AAW28928	S. avidini protein
42	135.5	16.0	673	17	AAW04208	Streptavidin/lucif
43	135	16.0	186	21	AAW84022	Amino acid sequenc
44	134.5	15.9	122	21	AAW80516	Streptomyces avidi
45	134.5	15.9	122	21	AAW80517	Streptomyces avidi

ALIGNMENTS

RESULT 1

AA44699

ID AAY44699 standard; Protein: 161 AA.

AC AAY44699;

XX 25-APR-2000 (first entry)

DT Potato proteinase inhibitor-I/Avidin fusion protein.

DE Potato proteinase inhibitor-I; pPI-I; streptavidin; worm; avidin;
KW insect; plant-noxious protein; pest resistance; moth; insect; weevil;
KW grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;
KW insecticidal; fusion protein.

XX Solanum tuberosum.

OS Synthetic.

XX Key Location/Qualifiers

FH Protein 1..31

FT /note= "Potato proteinase inhibitor-I sequence"

FT Peptide 1..23

FT /note= "Potato proteinase inhibitor-I signal peptide"

FT Protein 34..161

FT /note= "Avidin sequence"

WO200004049-A1.

XX 27-JAN-2000.

XX 15-JUL-1999; 99WO-NZ00110.

XX 15-JUL-1998; 98NZ-0331002.

DA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
 XX Christeller JT, Sutherland PW, Murray C, Markwick NP, Philip BA;
 PI Malone LA, Burgess EDJ;
 XX
 DR WPI: 2000-171244/15.
 DR N-PSDB: AA749865.
 XX
 PT New chimeric polypeptide and composition comprising the polypeptide
 PT useful for conferring pest resistance on plants -
 XX Example 2; Fig 8; 11pp; English.
 XX
 CC The present sequence is the potato proteinase inhibitor-I (PPI-I)/Avidin
 CC fusion protein. This comprises Avidin mature peptide, a plant-noxious
 CC protein, fused to the PPI-I signal peptide. Transformation of plant
 CC genome with the vector encoding the fusion protein can produce pest
 CC resistance in plants, plant derived products and stored harvest
 CC material. Pests that can be controlled include, cotton bollworm,
 CC tropical army-worm, European corn-borer or red mite, tobacco horn worm,
 CC loopers, rice stem borer, porina, cutworms, diamondback moth, potato
 CC tuber moth, codling moth, Indian meal moth, gypsy moth, argentine stem
 CC weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat
 CC weevils, mealworms, flour beetles, black field cricket, locusts,
 CC sawflies, Western flower thrips, Hessian flies or two-spotted mite.
 XX
 SQ Sequence 161 AA;

Query Match 100.0%; Score 845; DB 21; Length 161;
 Best Local Similarity 100.0%; Pred. No. 3.6e-83;
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESKEAHIVFFLLATPEFELLARKESDGEIPARKCSLTGKWTNDLGSNMITGAVNSRG 60
 Db 1 meskfahivffllatpelfellarkesdgpelparkcsltgkwtndlgsnmilgavnsrg 60

QY 61 EFTGTITAVTATSNKEIKESPLHGHTQNTINKRTQPTFGFTVNMKFSSESTVFTQCCFIDR 120
 Db 61 efgtyitavtateneikesplhgtqntinkrtqptfgftvnmkfssestvtftqccfidr 120

QY 121 NGREVLKTMWLLRSSVNDIGDDWKATRVGINIFTRLRTQKE 161
 Db 121 ngrevlktmwllrssvndigddwkatrvginiftrlrtqke 161

RESULT 2
 AAW29307
 ID AAW29307 standard; Protein: 152 AA.
 AC
 XX AAW29307;
 DT 27-APR-1998 (first entry)
 XX
 DE Wild-type avidin protein.
 XX
 KW Streptavidin; biotin; anti-interference reagent; detection; mutein;
 KW avidin; non-specific binding.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT /label= signal
 FT Protein 25..152
 XX
 PN DE19637718-A1.
 XX
 PD 02-OCT-1997.
 XX
 PF 16-SEP-1996; 96DE-1037718.
 XX
 PR 01-APR-1996; 96DE-1013053.

XX (BOEF) BOEHRINGER MANNHEIM GMBH.
 XX Brandtetter H, Deger A, Engh R, Kopetzki E, Mueller R;
 PI Schmitt U;
 XX
 DR WPI: 1997-492043/45.
 DR N-PSDB: AAT73194.
 XX
 PT Streptavidin and avidin muteins with reduced binding affinity for
 PT biotin - useful for reducing interference from nonspecific binding
 PT in assays
 XX
 PS Disclosure; Page 19-20; 26pp; German.
 XX
 CC This sequence represents an avidin which is used in a novel method
 CC of reducing interference from non-specific binding in assays. Muteins
 CC constructed from a core streptavidin or avidin sequence are selected that
 CC differ from the native polypeptide by at least one amino acid and have a
 CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable
 CC polypeptide may be present as a polymeric conjugate, e.g. with another
 CC polypeptide or protein, especially bovine serum albumin. These muteins
 CC are used as anti-interference reagents for reducing and/or avoiding
 CC nonspecific interactions in a process for detecting an analyte. In
 CC particular, they are used in assays where the streptavidin/avidin-biotin
 CC specific binding pair is involved for qualitative and/or quantitative
 CC determination of an analyte in a test sample, e.g. a heterogeneous
 CC immunoassay or a hybridisation assay. Despite having a lower binding
 CC affinity for biotin, the muteins have high immunological cross-reactivity
 CC with native streptavidin and avidin.
 XX
 SQ Sequence 152 AA;

Query Match 80.9%; Score 684; DB 18; Length 152;
 Best Local Similarity 97.7%; Pred. No. 7.6e-66;
 Matches 129; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 30 PEIPARKCSLTGKWTNDLGSNMITGAVNSRGFTQVITAVTATSNKEIKESPLHGHTQNTI 89
 Db 21 pglarkcsltgkwtndlgsnmilgavnsrgftqyltavlatsnelkesplhgtqnti 80

QY 90 NKRTQPTFGFTVNMKFSSESTVFTQCCFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
 Db 81 nkrtqptfgftvnmkfssestvtftqccfidrngkevlktmwllrssvndigddwkatrvg 140

QY 150 INIFTRLRTQKE 161
 Db 141 iniftrlrtqke 152

RESULT 3
 AAY17867
 ID AAY17867 standard; Protein: 152 AA.
 AC AAY17867;
 XX
 XX 20-AUG-1999 (first entry)
 DT
 XX Chicken avidin.
 DE
 XX Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;
 KW hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin.
 XX
 OS Gallus gallus.
 XX
 PN WO9929838-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 09-DEC-1998; 98WO-US26086.
 XX
 PR 09-DEC-1997; 97US-0067978.

```

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Cederholm-Williams SA;
XX WPI: 1999-385599/32.
XX N-PSDB; AAX80197.
XX A fibrinogen-converting enzyme fusion protein
XX Disclosure; Page 27; 35pp; English.
XX
XX The present invention describes a fibrinogen-converting enzyme fusion
XX protein (FCE). The fusion protein is a multidomain protein comprising:
XX (a) a FCE; and (b) a first member of a binding pair, that is linked to
XX the FCE chain; (i) directly by bonds utilizing the N-terminal amino
XX groups, the C-terminal carboxy groups or side-chain functionalities;
XX (ii) via a bifunctional linkage moiety linking the groups or
XX functionalities; or (iii) by the first member binding to the second
XX member of the binding pair, where the second member of the binding pair
XX is covalently attached to the first polypeptide chain. The FCE can be
XX used in a method for producing fibrin. Fibrin is useful as a sealant in
XX surgery to, e.g. reduce bleeding by sealing blood vessels, and tissues
XX that have been dissected either in surgery or through wounding. The
XX fusion protein allows for the removal of the fibrinogen converting
XX enzyme from the fibrin sealant preparation via the binding of
XX streptavidin to a biotin solid support. The present sequence represents
XX chicken (Gallus gallus) avidin as given in the present invention.
XX
XX Sequence 152 AA;
XX
XX Query Match 80.9%; Score 684; DB 20; Length 152;
XX Best Local Similarity 97.7%; Pred. No. 7.6e-66;
XX Matches 129; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 30 PEIPARKSLTGKWTNDLGSNMFTIGAVNSRGFTGTGYITAVTATSNKEIKESPLHGTONTI 89
XX I : |||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 21 pglarkcsltgkwtndlgsnmftigavnsrgftgtgyitavtatsneikesplhgtpnti 80
XX
XX QY 90 NKRTQPTFGFTVNMKFSESTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
XX I : |||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 81 nkrtqptfgftvnmkfsestvtftgqcfidrngkevlktmllrsvndigddwkatrvg 140
XX
XX QY 150 INIFTRLRTQKE 161
XX I : |||||||||||
XX Db 141 iniftrlrrtqke 152
XX
XX RESULT 4
XX AAY30130
XX ID AAY30130 standard; peptide; 152 AA.
XX AC AAY30130;
XX DT 26-OCT-1999 (first entry)
XX DE Amino acid sequence of the avidin monomer.
XX KW Avidin; biotin; neutravidin; ligand-binding molecule;
XX KW cross-linked structure; polynucleotide delivery; DNA linking;
XX KW biotin-avidin networked gene system; BANG system; DNA vaccine.
XX OS Unidentified.
XX
XX Key Location/Qualifiers
XX FH Peptide 1..24
XX FT Disulfide-bond 28..107
XX FT Modified-site 41
XX FT /note= "linked to a carbohydrate chain"
XX FT Misc-difference 58
XX FT /note= "this residue may also be Thr"

```

```

FT Misc-difference 57
FT /note= "biotin binding occurs here"
XX WO9939744-A1.
XX 12-AUG-1999.
XX 10-FEB-1999; 99WO-US02673.
XX 10-FEB-1998; 98US-0074213.
XX (OHIS ) UNIV OHIO STATE RES FOUND.
XX Luo D, Muller MT;
XX WPI: 1999-518369/43.
XX New cross-linked polynucleotide complexes, useful for cell-targeted
XX polynucleotide delivery
XX Example 1; Page 40; 135pp; English.
XX
XX The present sequence represents an avidin monomer. Avidin exists as a
XX tetramer, where each of the four monomer units binds to one biotin.
XX Biotin is complexed with a neutral avidin (neutravidin), and then
XX covalently coupled with polynucleotides, to exemplify the invention. The
XX specification describes a composition of complexes of polynucleotide
XX molecules covalently coupled to ligand moieties that are specifically
XX bound to ligand-binding molecules to form a cross-linked structure.
XX The composition allows incorporation of diverse oligonucleotides or
XX polynucleotides into a single complex for concomitant delivery into the
XX same cell. The composition is used to deliver polynucleotides to viable
XX cells. The method and compositions provide a new way of linking DNA
XX molecules and are useful for gene over-expression and non-covalent
XX cloning. The new system is called biotin-avidin networked gene (BANG)
XX system. It is possible that the BANG system can be used to link multiple
XX gene complexes to elicit broader immune reactions, e.g. as DNA vaccines.
XX The BANG system can also be used as a cloning tool. The BANG system also
XX eliminates the reading frame shift and cloning problems associated with
XX plasmids and vectors.
XX
XX Sequence 152 AA;
XX
XX Query Match 80.6%; Score 681; DB 20; Length 152;
XX Best Local Similarity 97.0%; Pred. No. 1.6e-65;
XX Matches 128; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 30 PEIPARKSLTGKWTNDLGSNMFTIGAVNSRGFTGTGYITAVTATSNKEIKESPLHGTONTI 89
XX I : |||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 21 pglarkcsltgkwtndlgsnmftigavnsrgftgtgyitavtatsneikesplhgnti 80
XX
XX QY 90 NKRTQPTFGFTVNMKFSESTVFTGQCIFDRNGKEVLKTMWLLRSSVNDIGDDWKATRVG 149
XX I : |||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX Db 81 nkrtqptfgftvnmkfsestvtftgqcfidrngkevlktmllrsvndigddwkatrvg 140
XX
XX QY 150 INIFTRLRTQKE 161
XX I : |||||||||||
XX Db 141 iniftrlrrtqke 152
XX
XX RESULT 5
XX AAY03098
XX ID AAY03098 standard; Protein; 400 AA.
XX AC AAY03098;
XX XX
XX DT 03-DEC-1999 (first entry)
XX DE Bovine scavenger receptor class A (ScR)/avidin fusion protein.
XX KW Scavenger receptor class A; ScR; avidin; fusion protein; bovine; ECD;
XX KW membrane-spanning domain; extracellular domain; biotin-binding activity;

```

KW endocytosis.
 XX
 OS Synthetic.
 OS BOS Taurus.
 XX
 PN WO9942577-A2.
 XX
 DD 26-AUG-1999.
 XX
 PF 23-FEB-1999; 99WO-GB00546.
 XX
 PR 23-FEB-1998; 98GB-0003757.
 PR 24-JUN-1998; 98GB-0013653.
 XX
 PA (EURO-) EUROGENE LTD.
 XX
 PI Via-Harttuala S. Kulomaa M. Lehtolainen P. Marjomaki V. Alronne K;
 XX
 DR WPI; 1999-561345/47.
 DR N-PSDB; AAZ09997).
 XX
 PF New fusion proteins having an extracellular domain with biotin-binding
 XX activity, used to target biotinylated molecules to specific sites in
 XX tissues -
 XX
 PS Claim 5; Page 21-23; 23pp; English.
 XX
 CC This invention describes a novel protein (A) which comprises a
 CC membrane-spanning domain and an extracellular domain (ECD), where the
 CC ECD comprises biotin-binding activity. Using the proteins or encoding
 CC nucleic acid molecules it is possible to target biotinylated molecules
 CC to specific sites in tissues. Molecules targeted in this way may be
 CC taken up by the tissues or cells by endocytosis, allowing the molecules
 CC to exert their effects within or on the cell. This sequence represents a
 CC bovine scavenger receptor class A/avidin fusion protein which is used in
 CC the description of the invention.
 XX
 SO Sequence 400 AA;
 }
 Query Match 80.2%; Score 678; DB 20; Length 400;
 Best Local Similarity 100.0%; Pred. No. 1.2e-64;
 Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 34 ARKCSLCKWNTDLSNMFTIGAVNSRGFTGTQVITAVTATSNKESPLHGTONTINKRT 93
 DD 273 ARKCSLTKWNTDLSNMFTIGAVNSRGFTGTQVITAVTATSNKESPLHGTONTINKRT 332
 QY 94 QPTFGFTVNMKFSSTVFTGQCFFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVGINIF 153
 Db 333 qptfgftvnmkfsstvtftgqcfidrngkevlktmllrsvndigddwkatrvginif 392
 QY 154 TRLRTOKE 161
 Db 393 trlrrtqke 400
 RESULT 6
 AAB35269
 ID AAB35269 standard; protein; 128 AA.
 XX
 AC AAB35269;
 XX
 DT 08-MAY-2001 (first entry)
 XX
 DE Chicken egg white avidin.
 XX
 KW Chicken; egg white; avidin; bacterial streptavidin; mutant; mutein;
 KW dimer; biotin binding activity; protein purification; immunoassay;
 KW signal transduction.
 XX
 OS Gallus gallus.
 OS
 XX
 PN WO200105977-A1.
 XX
 PD 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000WO-US18959.
 XX
 PR 15-JUL-1999; 99US-0354097.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Kulomaa MS, Bayer EA, Wilchek M, Laitinen OH;
 XX
 DR WPI; 2001-147344/15.
 XX
 PF Avidin and streptavidin mutants comprising tryptophan 110 and
 XX tryptophan 120 respectively, substituted by lysine for use in all known
 XX PT avidin/streptavidin-biotin systems and a purification process -
 XX
 PS Disclosure; Page 43; 49pp; English.
 XX
 CC The present invention provides a mutant chicken egg white avidin protein
 CC and a mutant bacterial streptavidin, which are capable of generating
 CC stable dimers with reversible biotin binding activity. These can be used
 CC in signal transduction studies, protein purification and immunoassays.
 CC The present sequence is the chicken egg white avidin protein.
 XX
 SO Sequence 128 AA;
 }
 Query Match 79.3%; Score 670; DB 22; Length 128;
 Best Local Similarity 98.4%; Pred. No. 2e-64;
 Matches 126; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 34 ARKCSLTGKWTNDLSNMFTIGAVNSRGFTGTVITATSNKESPLHGTONTINKRT 93
 Db 1 arkcsltgkwtndlsgnmftigavnsrgftgtvttavtatsnkesplhgtontinkrt 60
 QY 94 QPTFGFTVNMKFSSTVFTGQCFFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVGINIF 153
 Db 61 qptfgftvnmkfsstvtftgqcfidrngkevlktmllrsvndigddwkatrvginif 120
 QY 154 TRLRTOKE 161
 Db 121 trlrrtqke 128
 RESULT 7
 AAB35271
 ID AAB35271 standard; protein; 128 AA.
 XX
 AC AAB35271;
 XX
 DT 08-MAY-2001 (first entry)
 XX
 DE Chicken egg white avidin W10K mutant.
 XX
 KW Chicken; egg white; avidin; bacterial streptavidin; mutant; mutein;
 KW dimer; biotin binding activity; protein purification; immunoassay;
 KW signal transduction.
 XX
 OS Gallus gallus.
 OS Synthetic.
 XX
 PN WO200105977-A1.
 XX
 PD 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000WO-US18959.
 XX
 PR 15-JUL-1999; 99US-0354097.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 XX

PN WO200105977-A1.
 XX
 PD 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000WO-US18959.
 XX
 PR 15-JUL-1999; 99US-0354097.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Kulomaa MS, Bayer EA, Wilchek M, Laitinen OH;
 XX
 DR WPI; 2001-147344/15.
 XX
 PF Avidin and streptavidin mutants comprising tryptophan 110 and
 XX tryptophan 120 respectively, substituted by lysine for use in all known
 XX PT avidin/streptavidin-biotin systems and a purification process -
 XX
 PS Disclosure; Page 43; 49pp; English.
 XX
 CC The present invention provides a mutant chicken egg white avidin protein
 CC and a mutant bacterial streptavidin, which are capable of generating
 CC stable dimers with reversible biotin binding activity. These can be used
 CC in signal transduction studies, protein purification and immunoassays.
 CC The present sequence is the chicken egg white avidin protein.
 XX
 SO Sequence 128 AA;
 }
 Query Match 79.3%; Score 670; DB 22; Length 128;
 Best Local Similarity 98.4%; Pred. No. 2e-64;
 Matches 126; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 34 ARKCSLTGKWTNDLSNMFTIGAVNSRGFTGTVITATSNKESPLHGTONTINKRT 93
 Db 1 arkcsltgkwtndlsgnmftigavnsrgftgtvttavtatsnkesplhgtontinkrt 60
 QY 94 QPTFGFTVNMKFSSTVFTGQCFFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVGINIF 153
 Db 61 qptfgftvnmkfsstvtftgqcfidrngkevlktmllrsvndigddwkatrvginif 120
 QY 154 TRLRTOKE 161
 Db 121 trlrrtqke 128
 RESULT 7
 AAB35271
 ID AAB35271 standard; protein; 128 AA.
 XX
 AC AAB35271;
 XX
 DT 08-MAY-2001 (first entry)
 XX
 DE Chicken egg white avidin W10K mutant.
 XX
 KW Chicken; egg white; avidin; bacterial streptavidin; mutant; mutein;
 KW dimer; biotin binding activity; protein purification; immunoassay;
 KW signal transduction.
 XX
 OS Gallus gallus.
 OS Synthetic.
 XX
 PN WO200105977-A1.
 XX
 PD 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000WO-US18959.
 XX
 PR 15-JUL-1999; 99US-0354097.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 XX

PI Kulomaa MS, Bayer EA, Wilchek M, Laitinen OH;

DR WPI; 2001-147344/15.

PT Avidin and streptavidin mutants comprising tryptophan 110 and
 PT tryptophan 120 respectively, substituted by lysine for use in all known
 PT avidin/streptavidin-biotin systems and a purification process -

XX Disclosure; Page 45; 49pp; English.

PS The present invention provides a mutant chicken egg white avidin protein
 CC and a mutant bacterial streptavidin, which are capable of generating
 CC stable dimers with reversible biotin binding activity. These can be used
 CC in signal transduction studies, protein purification and immunoassays.
 CC The present sequence is the mutant chicken egg white avidin protein.

XX Sequence 128 AA;

Query Match 77.6%; Score 656; DB 22; Length 128;
 Best Local Similarity 97.7%; Pred. No. 6.3e-63;
 Matches 125; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 34 ARKSLTGKWTNDLGSNMTIGAVNSRGFTGYITAVTATSNEIKESPLHGQTONTINKRT 93
 Db 1 arkcsltgkwtndlgsnmtlgavnsrgftgyttavtatsneikesplhgtentinkrt 60
 Qy 94 QPTGFTVNWKSESTVFTGQCFFIDRNGKEVLKTMWLLRSSVNDIGDDWKATRVGINIF 153
 Db 61 qptgftvnwksestvtfgqcfidrngkevlktnmllrsvvndigddkkatrvginif 120
 Qy 154 TRLRTOKE 161
 Db 121 trlrrtqke 128

RESULT 8

AAR56485
 ID AAR56485 standard; Protein; 402 AA.

XX AC AAR56485;

XX DT 26-MAR-1995 (first entry)

XX DE ScFv PRAS110 and PRAS114.

XX KW Amplification; single chain variable region fusion protein; PCR.

XX OS Synthetic.

XX PN WO9415644-A.

XX PD 21-JUL-1994.

XX PF 17-JAN-1994; 94WO-GB00087.

XX PR 15-JAN-1993; 93GB-0000686.

XX PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

XX PI Deonarain M, Epenetos AA, Spooner RA;

XX DR WPI; 1994-248907/30.

DR N-PSDB; AAQ70661.

XX New cpds. comprising a targeting portion and a cytotoxic portion
 PT - used esp. for treating mammals for destroying target cells,
 PT partic. tumour cells

PS Disclosure; Fig 11; 114pp; English.

XX The sequence is that of the ScFv PRAS110 and PRAS114 between HindIII and
 CC EcoRI site obtd. by PCR.

CC See also AAR56482-4.

XX Sequence 402 AA;

Query Match 17.5%; Score 148; DB 15; Length 402;
 Best Local Similarity 34.1%; Pred. No. 1.8e-07;
 Matches 47; Conservative 18; Mismatches 57; Indels 16; Gaps 8;

Qy 29 GPEIPA---RKCSLTGKWTNDLGSNMTIGAVNSRGFTGYITAV-TATSNEIKESPLHG 84
 Db 267 gleapaapaegaigtgwynglsgflvta-gadgaltgyesavgnaesryvitgryds 325
 Qy 85 TQNTINKRTQPTGFTVNWK----FSSESTVFTGQCFFIDRNGKEV-LKTMWLLRSSVNDI 139
 Db 326 apatdgsrt--algwtvawknnrynnahtsattsggyv---ggaaearintqltsgtte- 379

Qy 140 GDDWKATRVGINIFRLR 157

Db 380 anawkstlvghdttfkvk 397

RESULT 9

AAR59217

ID AAR59217 standard; Protein; 183 AA.

XX AC AAR59217;

XX DT 27-AUG-1998 (first entry)

XX DE S. avidinii streptavidin mutant protein #1.

XX KW Streptavidin; ligand; binding affinity; mutant; isolation;

XX KW purification; recover; immobilise.

XX OS Streptomyces avidinii.

XX OS Synthetic.

XX PH Key Location/Qualifiers

FT Misc-difference 68..71

FT /label= VTAR68ESAV

FT /note= "Wild type ESAB is replaced by VTAR. Numbering
 is from the start of the mature protein"

XX PN EP835934-A2.

XX PD 15-APR-1998.

XX PF 09-OCT-1997; 97EP-0117504.

XX PR 10-OCT-1996; 96DE-1041876.

XX PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX PI Skerra A, Voss S;

XX DR WPI; 1998-218868/20.

DR N-PSDB; AAV34715.

XX Streptavidin mutants with higher binding affinity for peptide

PT ligands - have mutation in amino acid region 44-53, used to isolate,

PT purify or determine fusion proteins including these ligands

XX Disclosure; Page -; 21pp; German.

XX This sequence represents a mutant streptavidin protein isolated from
 CC Streptomyces avidinii where the residues ESAB at position 44-47 of
 CC the mature wild type sequence are replaced by VTAR. This sequence is
 CC used to produce mutants which are used in a method to assay the binding
 CC affinity of streptavidin mutants. These mutants have a mutation within
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
 CC binding affinity than the wild-type for peptide ligands that include
 CC the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa, Y

CC and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips. CC NOTE: This sequence does not appear in the specification but has been constructed from the wild-type streptavidin sequence represented in AAV34714.

XX Sequence 183 AA;

Query Match 17.2%; Score 145.5; DB 19; Length 183;
Best Local Similarity 32.3%; Pred. NO. 1.1e-07;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

QY 31 EIPARKSLCKWTNDLGSNMFTICAVNSRGFTGTITAV-TATSNKEIKESPLHGTQNTI 89
DB 33 QVSAEAGILGTWYNGIGSTFLVTA-GADGALTGYTARGNAESRYVLTGRYDSEAPAD 91
QY 90 NKRTQPTFGFTVNNK----FSESTVTFTGQCFIDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
DB 92 gsqt--algwtvawknnrynahsattwsqgyv---ggaeartntqwltsqtte-anawk 145
QY 145 ATRVGINIFTRLR 157
DB 146 stlvghdtftkvvk 158

RESULT 10
AAB35270
ID AAB35270 standard; protein; 159 AA.

XX AAB35270;
DT 08-MAY-2001 (first entry)

XX Streptomyces avidinii streptavidin.

XX Chicken; egg white; avidin; bacterial streptavidin; mutant; mutein;
KW dimer; biotin binding activity; protein purification; immunoassay;
KW signal transduction.

XX Streptomyces avidinii.

XX WO200105977-A1.

XX 25-JAN-2001.

XX 13-JUL-2000; 2000WO-US18959.

XX 15-JUL-1999; 99US-0354097.

XX (VEDA) VEDA RES & DEV CO LTD.

XX Kulomaa MS, Bayer EA, Wilchek M, Laitinen OH;

XX WPI; 2001-147344/15.

XX Avidin and streptavidin mutants comprising tryptophan 110 and
PT tryptophan 120 respectively, substituted by lysine for use in all known
PT avidin/streptavidin-biotin systems and a purification process -

XX Disclosure; Page 44; 49pp; English.

XX The present invention provides a mutant chicken egg white avidin protein
CC and a mutant bacterial streptavidin, which are capable of generating
CC stable dimers with reversible biotin binding activity. These can be used
CC in signal transduction studies, protein purification and immunoassays.
CC The present sequence is the Streptomyces avidinii streptavidin protein.

XX Sequence 159 AA;

Query Match 16.9%; Score 142.5; DB 22; Length 159;
Best Local Similarity 32.3%; Pred. NO. 2e-07;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

QY 31 EIPARKSLCKWTNDLGSNMFTICAVNSRGFTGTITAV-TATSNKEIKESPLHGTQNTI 89
DB 9 QVSAEAGILGTWYNGIGSTFLVTA-GADGALTGYTARGNAESRYVLTGRYDSEAPAD 67
QY 90 NKRTQPTFGFTVNNK----FSESTVTFTGQCFIDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
DB 68 gsqt--algwtvawknnrynahsattwsqgyv---ggaeartntqwltsqtte-anawk 121
QY 145 ATRVGINIFTRLR 157
DB 122 stlvghdtftkvvk 134

RESULT 11
AAP93531
ID AAP93531 standard; protein; 160 AA.

XX AAP93531;

XX 04-JUN-1990 (first entry)

XX Mature streptavidin as encoded by a synthetic gene.

XX Streptavidin; synthetic gene; Streptomyces avidinii;

XX Biotin; chimeric gene.

XX WO8903422-A.

XX 20-APR-1989.

XX 07-OCT-1988; 88WO-GB00831.

XX 08-OCT-1987; 87GB-0023661.

XX (BRBI-) BRIT BIO-TECHN LTD.

XX Edwards RW;

XX WPI; 1989-130040/17.

XX N-PSDB; AAN90755.

XX DNA sequence encoding streptavidin and vector -
PT comprising hybrid gene encoding fusion protein with
PT biotin-binding activity

XX Fig 2; page 2/5; 24pp; English.

XX The synthetic gene is constructed from 24 oligomers (BB214-BB237)
CC which were synthesised by automated phosphoramidite chemistry.
CC Apart from Met-1, it has the same sequence as natural mature streptavidin
CC protein produced by Streptomyces avidinii.

XX Sequence 160 AA;

Query Match 16.9%; Score 142.5; DB 10; Length 160;
Best Local Similarity 32.3%; Pred. NO. 2e-07;
Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps 7;

QY 31 EIPARKSLCKWTNDLGSNMFTICAVNSRGFTGTITAV-TATSNKEIKESPLHGTQNTI 89
DB 10 QVSAEAGILGTWYNGIGSTFLVTA-GADGALTGYTARGNAESRYVLTGRYDSEAPAD 68
QY 90 NKRTQPTFGFTVNNK----FSESTVTFTGQCFIDRNGKEV-LKTMWLLRSSVNDIGDDWK 144
DB 69 gsqt--algwtvawknnrynahsattwsqgyv---ggaeartntqwltsqtte-anawk 122
QY 145 ATRVGINIFTRLR 157


```

RESULT 14
AAP80160
ID AAP80160 standard; protein; 163 AA.
AC
XX AAP80160;
XX
XX
DT 13-OCT-1990 (first entry)
XX
XX Biosynthetic protein with streptavidin trailer.
DE
XX Biosynthetic Antibody Binding site (BABS): variable domain;
KW anti-CEA monoclonal antibody; streptavidin.
XX
XX Synthetic.
OS
XX
XX WO8809344-A.
PN
XX
XX 01-DEC-1988.
PD
XX
XX 19-MAY-1988; 88WO-US01737.
PF
XX
XX 21-MAY-1987; 87US-0052800.
PR
XX
XX (CREA-) CREATIVE BIOMOLECULES INC.
PA
XX
XX Huston JS, Oppermann H;
PI
XX
XX WFI; 1986-353926/49.
DR
XX
XX N-PSDB; AAN80186.
DT
XX
XX
XX Recombinant multifunctional protein -
PT having an antibody binding site and a sequence for biological activity,
PP ion sequestering or binding to a solid support,
XX
XX Disclosure; ; 115pp; English.
PS
XX
XX Multi-functional biosynthetic protein comprising single
CC Chain BABS and streptavidin protein trailer linked via a spacer
CU sequence.
CC See also AAN80171-N80192.
XX
XX
XX Sequence 163 AA;
XX

```

```

XX Streptomyces avidinl1.
XX
XX WO8705026-A.
XX
XX 27-AUG-1987.
XX
XX 24-FEB-1987; 87WO-US00397.
XX
XX 24-FEB-1986; 86US-0833324.
XX
XX (UWCO-) COLUMBIA UNIV N Y.
XX
XX (UYNX-) UNIV NEW YORK.
XX
XX Cantor CR, Axel R, Garana C;
XX
XX WPI; 1987-250198/35.
XX
XX N-PSDB; AAN70810.
XX
XX
XX DNA encoding Streptavidin - obtd. by restriction endo-nuclease
XX digestion of chromosomal DNA of Streptomyces avidinl1
XX
XX Disclosure; Figure 3; 54pp; English.
XX
XX The sequence is that of streptavidin from S.avidinl1. Streptavidin
XX may be expressed from a fusion gene comprising its coding gene and DNA
XX encoding a target protein of interest, where the streptavidin has
XX binding sites for biotin or deriv. The streptavidin has 4 binding
XX sites free for biotin, and is produced free of biotin
XX contamination. Improved streptavidins may also be produced by
XX site-directed mutagenesis. The fused gene may be used to produce
XX labelled, chemically-modified proteins in vivo, and to isolate
XX proteins when only the sequence of the gene is known.
XX
XX Sequence 182 AA;
XX
XX
XX Query Match 16.9%; Score 142.5; DB 8; Length 182;
XX Best Local Similarity 32.3%; Pred. No. 2.4e-07;
XX Matches 43; Conservative 20; Mismatches 57; Indels 13; Gaps
XX
XX QY 31 EIPARKSLGKKTNDLGSNNHTGAVNSRGFTGTYITAV-TATSNDEIKESPLHCTONTI 89
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 32 qvsaaesgitgwyndlgstfvta-gadgaltgtvesavgnadestvyltgydsapad 90
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 90 NKRTPTFGTIVNWK-----FSESTIVFTGTCFIDRNGKEY-LKTMWLLRSYNDIGDDWK 144
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 91 gsgt--algwcvakwnyrnahsattwsgyv---ggaaearintqlltsgtte-anawk 144
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 145 ATRVGINIFTRLR 157
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DB 145 stlvghdtfcvk 157
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX Search completed: September 17, 2002. 15:42:04
XX Job time: 265 sec

```

Search completed: September 17, 2002, 15:42:04
Job time: 265 sec